PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/82, 15/29, 15/57, 15/25, 15/15, 15/62, A01H 5/00

A1

(11) International Publication Number: WO 96/21029

(43) International Publication Date: 11 July 1996 (11.07.96)

(21) International Application Number: PCT/CA95/00724

(22) International Filing Date: 21 December 1995 (21.12.95)

(30) Priority Data:
08/366,783 30 December 1994 (30.12.94) US

(71) Applicant: UNIVERSITY TECHNOLOGIES INTERNATIONAL, INC. [CA/CA]; HM 382, 3330 Hospital Drive N.W., Calgary, Alberta T2N 4N1 (CA).

(72) Inventor: MOLONEY, Maurice; 131 Edgehill Place, Calgary, Alberta T3A 2S4 (CA).

(74) Agent: GRAVELLE, Micheline; Smart & Biggar, Suite 2300, 439 University Avenue, Toronto, Ontario M5G 1Y8 (CA).

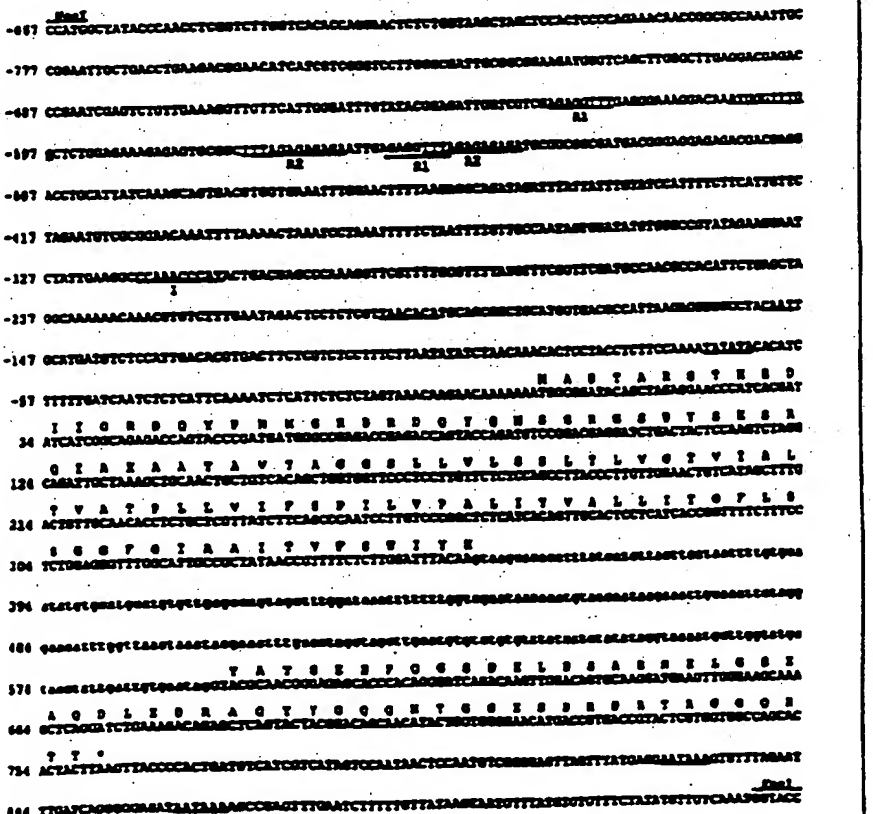
(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: OIL BODY PROTEINS AS CARRIERS OF HIGH VALUE PROTEINS

(57) Abstract

The present invention relates to the use of a class of genes called oil body protein genes that have unique features. The discovery of these features allowed the invention of methods for the production of recombinant proteins wherein a protein of interest can be easily separated from other host cell components. The invention is further exemplified by methods for exploitation of the unique characteristics of the oil body proteins and oil body genes for expression of polypeptides of interest in many organisms, particularly plant seeds. Said polypeptides may include but are not limited to: seed storage proteins, enzymes, bioactive peptides, antibodies and the like. The invention can also be modified to recover recombinant polypeptides fused to oleosins from non-plant host cells. Additionally the invention provides a method of using recombinant proteins associated with seed oil bodies released during seed germination for expression of polypeptides that afford protection to seedlings from pathogens. Finally, the persistent association of oil body proteins with the oil body can be further utilized to develop a biological means to create novel immobilized enzymes useful for bioconversion of substrates.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia
AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

OIL BODY PROTEINS AS CARRIERS OF HIGH VALUE PROTEINS

Field of the Invention

The present application relates to a method of producing polypeptides in plants.

Background of the Invention

Many very diverse methods have been tested for the production of recombinant molecules of interest and commercial value. Different organisms that have been considered as hosts for foreign protein expression include singled celled organisms such as bacteria and yeasts, cells and cell cultures of animals, fungi and plants and whole organisms such as plants, insects and transgenic animals.

Plants represent a highly effective and economical means to produce recombinant proteins as they can be grown on a large scale with modest cost inputs and most commercially important species can now be transformed. Although the expression of foreign proteins has been clearly demonstrated the development of systems with commercially viable levels of expression coupled with cost effective separation techniques has been limited.

The present inventor has developed a method of producing recombinant proteins in plants which is described in PCT published application no. WO 93/21320 which is incorporated herein by reference.

Application no. WO 93/21320 describes the use of an oleosin gene to target the expression of a polypeptide to an oil body in a host cell. In particular, the method involved transforming a plant host cell with a chimeric DNA sequence comprising (i) a sufficient portion of an oleosin gene to provide targeting to an oil body and (ii) a DNA encoding the polypeptide of interest. The transformed plant cells are grown and the polypeptide of interest is expressed as a fusion protein with the oleosin protein in the oil bodies of the seed. In order to recover the polypeptide, the oil bodies are isolated from the seed and disrupted to release the polypeptide oleosin fusion protein. The polypeptide can then be cleaved from

oleosin. The unique features of both the oleosin protein and the oleosin expression patterns are used to provide a means of synthesizing commercially important proteins on a scale that is difficult if not impossible to achieve using conventional systems of protein production. The use of plants to produce proteins of interest allows exploitation of the ability of plants to capture energy and limited nutrient input to make proteins. The scale and yield of material afforded by production in plants allows adaptation of the technology for use in the production of a variety of polypeptides of commercial interest.

SUMMARY OF THE INVENTION

The present inventor has now developed useful improvements in, and new applications for, the method of producing recombinant polypeptides described in WO 93/21320.

In the broadest sense, the method of the present invention provides a method for the expression of a recombinant polypeptide by a host cell said method comprising: a) introducing into a host cell a chimeric DNA sequence comprising: 1) a first DNA sequence capable of regulating the transcription in said host cell of 2) a second DNA sequence, wherein said second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to a lipid phase linked in reading frame to (ii) a DNA sequence encoding said recombinant polypeptide; and 3) a third DNA sequence encoding a termination region functional in the host cell; and b) growing said host cell to produce the recombinant fusion polypeptide.

In one embodiment the recombinant polypeptide is enzyme. The processing of a wide variety of materials using enzymes has enormous commercial potential. The present invention provides for methods to produce recombinant enzymes in mass quantities which can be separated from cellular components by partitioning of the oil-body fraction. The enzyme of interest may be cleaved from the oleosin or may be used in association with the oil-body fraction. The enzyme, while still part

of the oleosin fusion polypeptide associated with the oil body, may retain its enzymatic properties. Enzymes fused to oleosins in an oil-body fraction represent a type of immobilized and reusable enzyme system. Immobilized enzyme systems have been developed in association with various inert support matrices for many industrial purposes including cellulose beads, plastic matrixes and other types of inert materials. Enzymes attached to oil-bodies can be mixed with solutions containing enzyme substrates and subsequently recovered by floatation and partitioning of the oil-body fraction and reused.

In another embodiment, the chimeric DNA encoding the oil body protein and the polypeptide further includes a linker DNA sequence encoding an amino acid sequence that is specifically cleavable by enzymatic or chemical means. This allows the polypeptide to be easily separated from the oleosin fusion by contacting the oil bodies with the appropriate enzyme or chemical.

Accordingly, the present invention provides a method for the expression and release of a recombinant polypeptide by a host cell said method comprising: a) introducing into a host cell a chimeric DNA sequence comprising: 1) a first DNA sequence capable of regulating the transcription in said host cell of 2) a second DNA sequence, wherein said second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to a lipid phase linked in reading frame to (ii) a DNA sequence encoding said recombinant polypeptide and (iii) a linker DNA sequence encoding an amino acid sequence that is specifically cleavable by enzymatic or chemical means wherein said linker DNA sequence (iii) is located between said DNA sequence (i) and (ii); and 3) a third DNA sequence encoding a termination region functional in the host cell; b) growing said host cell to produce the recombinant fusion polypeptide and c) contacting the lipid phase with said enzymatic or chemical means such that said recombinant polypeptide is released from the recombinant fusion polypeptide.

The invention thus provides methods for the separation of recombinant

proteins from host cell components by partitioning of the oil body fraction and subsequent release of the recombinant protein via specific cleavage of the recombinant protein - oleosin fusion. Optionally a cleavage site may be located prior to the N-terminus and after the C-terminus of the polypeptide of interest allowing the fusion polypeptide to be cleaved and separated by phase separation into its component peptides. This production system finds utility in the production of many proteins and peptides such as those with pharmaceutical, enzymic, rheological and adhesive properties.

In a specific embodiment of the above, the recombinant polypeptide is an enzyme. In particular, the enzyme may be specific for the amino acid sequence encoded by the linker DNA sequence (iii). In such a case, the enzyme can effectively auto-release by cleaving itself from the fusion protein.

Accordingly, the present invention yet also provides a method of preparing an enzyme in a host cell in association with an oil body and releasing said enzyme from the oil body, said method comprising: a) transforming a host cell with a chimeric DNA sequence comprising: 1) a first DNA sequence capable of regulating the transcription of 2) a second DNA sequence, wherein said second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to an oil body; (ii) a DNA sequence encoding an enzyme and (iii) a linker DNA sequence located between said DNA sequence (i) encoding the oil body protein gene and said DNA sequence (ii) encoding the enzyme and encoding an amino acid sequence that is cleavable by the enzyme encoded by the DNA sequence (ii); and 3) a third DNA sequence encoding a termination region functional in said host cell b) growing the host cell to produce the recombinant fusion polypeptide under conditions such that enzyme is not active; c) recovering the oil bodies containing the recombinant fusion polypeptide; and d) altering the environment of the oil bodies such that the enzyme is activated and cleaves itself from the recombinant fusion polypeptide.

In a further embodiment of the above, two different chimeric DNA sequences can be prepared and expressed in different host cells. One chimeric DNA sequence may contain a DNA sequence encoding the oil body protein linked to a DNA sequence encoding a desired polypeptide via the DNA linker encoding an amino acid sequence that is cleavable by enzymatic means. The second chimeric DNA sequence may contain a DNA sequence encoding the oil body protein linked to a DNA sequence encoding an enzyme that can cleave the amino acid sequence encoded by the linker in the first chimeric DNA sequence. When these chimeric DNA sequences are expressed as fusion proteins by transformed host cells and associated with the oil bodies, the two oil body fractions may be mixed so that the enzyme portion of the second protein fusion cleaves the polypeptide of the first protein fusion.

Accordingly, the present invention further provides a method for the expression of a recombinant polypeptide by a host cell in association with an oil body and separating said recombinant polypeptide from the oil body, said method comprising: a) transforming a first host cell with a first chimeric DNA sequence comprising: 1) a first DNA sequence capable of regulating the transcription in said host cell of 2) a second DNA sequence, wherein said second sequence encodes a first recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to a lipid phase linked in reading frame to (ii) a DNA sequence encoding said recombinant polypeptide; and (iii) a linker DNA sequence encoding an amino acid sequence that is specifically cleavable by enzymatic means wherein said linker DNA sequence (iii) is located between said (i) DNA sequence encoding the oil body protein and said (ii) DNA sequence encoding the recombinant polypeptide; and 3) a third DNA sequence encoding a termination region functional in the host cell; and b) transforming a second host cell with a second chimeric DNA sequence comprising: 1) a first DNA sequence capable of regulating the transcription specifically during seed germination and seed growth of

2) a second DNA sequence wherein said second sequence encodes a second recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the second recombinant fusion polypeptide to a lipid phase linked in reading frame to a DNA sequence, encoding a specific enzyme that is capable of cleaving the linker DNA sequence of said first chimeric DNA sequence; and 3) a third DNA sequence encoding a termination region; c) growing said first host cell under conditions such that the first recombinant fusion polypeptide is expressed and associated with the oil bodies to produce a first oil body fraction containing the first recombinant fusion polypeptide; d) growing said second host cell under conditions such that the second recombinant fusion polypeptide is expressed and associated with the oil bodies to product a second oil body fraction containing the second recombinant fusion polypeptide; e) contacting the first oil body fraction of step (c) with the second oil body fraction of step (d) under conditions such that the enzyme portion of the second recombinant fusion polypeptide cleaves the first recombinant polypeptide from the first recombinant fusion polypeptide.

In addition to the production and isolation of recombinant proteins from plants the present invention also contemplates methods for crop improvement and protection. The nutritional quality of seeds has been improved by the addition of proteins with high levels of essential amino acids (DeClercq et al., 1990, Plant Physiol. 94:970-979) and enzymes such as lauroyl-ACP thioesterase from *Umbellularia californica* that affect lipid composition (US Patent 5,298,421). To date these seed modifications have only been conducted using seed storage gene promoters that may have inherent limitations. Use of oleosin regulatory sequences provides an additional means by which to accomplish such modifications.

Insect predation and fungal diseases of crop plants represent two of the largest causes of yield losses. A number of strategies dependent on transformation and expression of recombinant proteins in plants have been advanced for the protection of plants from insects and fungi (Lamb et al., 1992, Bio/Technology

11:1436-1445). These strategies are exemplified by the expression of peptide inhibitors of insect digestive enzymes such as cowpea trypsin inhibitor (Hoffman et al., 1992, J. Economic Entomol. 85: 2516-1522) bacterial or arachnid protein toxins (Gordon and Zlotkin, 1993, FEBS Lett., 315:125-128) and the expression of chitinase enzymes for the digestion of fungal cell walls (Broglie et al., 1991, Science 254: 5035, 1194-1197; Benhamou et al., 1993, Plant Journal 2:295-305; Dunsmuir et al., 1993, *In* Advances in molecular genetics of plant-microbe interactions, Vol 2. pp 567-571, Nester, E.W. and Verma, D.P.S. eds.). The use of oleosin proteins to localize specific polypeptides that afford crop protection allows one to develop novel strategies to protect vulnerable germinating seeds.

The use of oleosins whose expression is limited to pollen allows one to alter the function of pollen to specifically control male fertility. One may use promoter sequences from such oleosins to specifically express recombinant proteins that will alter the function of pollen. One such example is the use of such promoters to control the expression of novel recognition proteins such as the self-incompatibility proteins. Additional uses are contemplated including expression of oleosin fusion proteins in pollen that are toxic to pollen. Seed specific oleosins may be used to alter female fertility.

Accordingly, the present invention also provides a method for the production and release of a recombinant polypeptide from a recombinant fusion polypeptide associated with a plant oil body fraction during seed germination and plant seedling growth, said method comprising: a) introducing into a plant cell a first chimeric DNA sequence comprising: 1) a first DNA sequence capable of regulating the transcription in said plant cell of 2) a second DNA sequence wherein said DNA second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to an oil body, linked in reading frame to (ii) a DNA sequence encoding a recombinant polypeptide and (iii) a linker DNA sequence encoding an amino acid sequence that is specifically cleavable by

enzymatic means wherein said linker DNA sequence (iii) is located between said DNA sequence (i) encoding the oil body protein and said DNA sequence (ii) encoding the recombinant polypeptide; and 3) a third DNA sequence encoding a termination region; b) sequentially or concomitantly introducing into the genome of said plant a second chimeric DNA sequence comprising: 1) a first DNA sequence capable of regulating the transcription specifically during seed germination and seed growth of 2) a second DNA sequence encoding a specific enzyme that is capable of cleaving the linker DNA sequence of said first chimeric DNA sequence; and 3) a third DNA sequence encoding a termination region; c) regenerating a plant from said plant cell and growing said plant to produce seed whereby said recombinant fusion polypeptide is expressed and associated with oil bodies and d) allowing said seed to germinate wherein said enzyme in said second chimeric DNA sequence is expressed and cleaves the recombinant polypeptide from the recombinant fusion polypeptide associated with the oil bodies during seed germination and early seedling growth.

The present invention further provides a method for producing an altered seed meal by producing a recombinant polypeptide in association with a plant seed oil body fraction, said method comprising: a) introducing into a plant cell a chimeric DNA sequence comprising: 1) a first DNA sequence capable of regulating the transcription in said plant cell of 2) a second DNA sequence wherein said second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to an oil body, linked in reading frame to (ii) a DNA sequence encoding a recombinant polypeptide and 3) a third DNA sequence encoding a termination region; b) regenerating a plant from said plant cell and growing said plant to produce seed whereby said recombinant polypeptide is expressed and associated with oil bodies; and c) crushing said seed and preparing an altered seed meal.

The present invention includes within its scope all of the above described

chimeric DNA sequences.

In the broadest sense, the present invention provides a chimeric DNA sequence, capable of being expressed in association with an oil body of a host cell comprising: 1) a first DNA sequence capable of regulating the transcription in said host cell of 2) a second DNA sequence, wherein said second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to a lipid phase linked in reading frame to (ii) a DNA sequence encoding said recombinant polypeptide; and 3) a third DNA sequence encoding a termination region functional in the host cell.

The present invention also includes within its scope a plant, plant cell or plant seed containing any of the chimeric DNA sequences of the present invention.

The methods described above are not limited to recombinant proteins produced in plant seeds as oleosins may also be found in association with oil bodies in other cells and tissues. Additionally the methods are not limited to the recovery of recombinant proteins produced in plants because the extraction and release methods can be adapted to accommodate oleosin protein fusions produced in any cell type or organism. An extract containing the oleosin recombinant protein fusion is mixed with additional oleosins and appropriate tri-glycerides and physical conditions are manipulated to reconstitute the oil-bodies. The reconstituted oil-bodies are separated by floatation and the recombinant proteins released by the cleavage of the junction with oleosin.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a schematic representation of the types of oil body protein fusions that are contemplated as methods of the invention for the fusion of oil-body protein genes with genes encoding foreign polypeptides. IA is a C-terminal fusion of a desired polypeptide to a oil body protein; IB is an N-terminal fusion of a desired polypeptide to oil body protein; IC is an internal fusion of a desired polypeptide within oil body protein; and ID is an inter-dimer translational fusion of

desired polypeptide enclosed between two substantially complete oil body protein targeting sequences. Each fusion is shown in a linear diagrammatic form and in the configuration predicted when specifically associated with the oil body. In both the linear and oil body associated form, the oil body coding sequence that specifically targets the protein to the oil body is shown as a single thin line, a solid circle represents a protease recognition motif; a corkscrew line represents a native C- or N-terminal of a oil body protein and a inserted coding region is represented by an open box. The oil body is represented as a simple circle.

Figure 2 shows the nucleotide sequence (SEQ ID NO.1) and deduced amino acid sequence (SEQ ID NO.2) of an oil-body protein gene that codes for a 18 KDa oleosin from *Arabidopsis thaliana*. The intron sequence is printed in lower case. The predicted amino acid sequence is shown in single letter code.

Figure 3 shows a schematic representation of the construction of pOleoP1.

Figure 4 shows the nucleotide sequence (SEQ ID NO.3) of a *B. napus* oleosin cDNA clone and the predicted amino acid sequence (SEQ ID NO.4).

Figure 5 describes the construction of a oleosin/GUS fusion for expression in *E. coli*.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

In accordance with the subject invention, methods and compositions are provided for a novel means of production of recombinant proteins and peptides that can be easily separated from host cell components. In accordance with further embodiments of the invention methods and compositions are provided for novel uses of recombinant proteins produced by said methods.

In accordance with one aspect of the subject invention, methods and compositions are provided for a novel means of production of recombinant proteins and peptides in host cells that are easily separated from other host cell components. Purification of the recombinant protein, if required, is greatly simplified. The recombinant DNA encoding the peptide of interest may be part or all of a naturally occurring gene from any source, it may be a synthetic DNA sequence or it may be

a combination of naturally occurring and synthetic sequences. The subject method includes the steps of preparing an expression cassette comprising a first DNA sequence capable of regulating the transcription of a second DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting to an oil body and fused to this second DNA sequence a third DNA sequence encoding the protein, polypeptide or RNA of interest; delivery and incorporation of the expression cassette into a host cell; production of a transformed organism or cell population in which the chimearic gene product is expressed and recovery of a chimearic gene protein product through specific association with an oil body. The peptide of interest is usually a foreign polypeptide normally not expressed in the host cell or found in association with the oil-body.

The transformed host cells may be from any source including plants, fungi, bacteria and animals. In a preferred embodiment the host cell is a plant and the chimeric product is expressed and translocated to the oil bodies of the seed.

The use of an oil body protein as a carrier or targeting vehicle provides a simple mechanism to recover recombinant proteins. The chimeric protein associated with the oil body or reconstituted oil body fraction is separated away from the bulk of cellular components in a single step (such as centrifugation or floatation); the protein is also protected from degradation during extraction as the separation also reduces contact of the recombinant proteins with non-specific proteases.

The invention contemplates the use of recombinant proteins, specifically enzymes, fused to oleosins and associated with oil bodies, or reconstituted oil bodies for conversion of substrates in aqueous solutions following mixing of oil body fractions and substrate solutions. Association of the recombinant enzyme with the oil body allows subsequent recovery of the recombinant enzyme by simple means (centrifugation and floatation) and repeated use thereafter.

In accordance with further embodiments of the invention methods and compositions are provided for the release of recombinant proteins and peptides

fused to oleosin proteins specifically associated with isolated oil body or reconstituted oil body fractions. The subject method includes the steps of preparing an expression cassette comprising a first DNA sequence capable of regulating the transcription of a second DNA sequence encoding a sufficient portion of an oil body protein gene such as oleosin to provide targeting to an oil body and fused to this second DNA sequence via a linker DNA sequence encoding a amino acid sequence cleavable by a specific protease or chemical treatment a third DNA sequence encoding the protein, polypeptide or RNA of interest; such that the protein of interest can be cleaved from the isolated oil body fraction by the action of said specific chemical or protease.

For embodiments of the invention wherein the cleavage of recombinant proteins fused to oleosins associated with seed oil bodies is contemplated in germinating seed the expression cassette containing said recombinant protein gene so described above is modified to contain an additional second recombinant DNA molecule comprising a first DNA sequence capable of regulating expression in plants, particularly in germinating seed, more specifically seed embryo or other seed tissue containing oil bodies and under the control of this regulatory sequence a DNA sequence encoding a protease enzyme, specifically a particular protease enzyme capable of cleavage of said recombinant chimeric proteins associated with said oil bodies to release a protein or peptide of interest from the oil body, and a transcriptional and translational termination region functional in plants. It is desirable that the second recombinant DNA molecule be so constructed such that the first and second recombinant DNA sequences are linked by a multiple cloning site to allow for the convenient substitution of any one of a variety of proteolytic enzymes that may be used to cleave chimeric proteins associated with oil bodies.

It is obvious to a person skilled in the art of plant molecular biology, genetics or plant breeding that the equivalent to the above modification to the expression cassette to allow release of proteins and peptides of interest in germinating seeds can be accomplished by other similar means. For example it is

possible that the first recombinant DNA molecule and the second recombinant DNA molecule described above may be contained within two independent expression cassettes introduced into the genome of a plant independently. Additionally it is possible to sexually cross a first recombinant plant containing the first recombinant DNA molecule integrated into its genome with a second recombinant plant with the second recombinant DNA integrated into its genome to produce seed comprising both the first and second recombinant DNA molecules.

For embodiments of the invention wherein the recombinant protein is to be produced in and potentially recovered from plant seeds the expression cassette will generally include, in the 5'-3' direction of transcription, a first recombinant DNA sequence comprising a transcriptional and translational regulatory region capable of expression in plants, particularly in developing seed, more specifically seed embryo or other seed tissue that has oil body or triglyceride storage such as pericarp or cuticle, and a second recombinant DNA sequence encoding a chimeric peptide or protein comprising a sufficient portion of an oil body specific protein to provide targeting to an oil body, a protein of interest, and a transcriptional and translational termination region functional in plants. One or more introns may also be present within the oil body specific protein coding sequence or within the coding sequence of the protein of interest. The chimeric peptide or protein may also comprise a peptide sequence linking the oil body specific portion and the peptide or protein of interest that can be specifically cleaved by chemical or enzymatic means. It is desirable that the DNA expression cassette be so constructed such that the first and second recombinant DNA sequences are linked by a multiple cloning site to allow for the convenient substitution of alternative second recombinant DNA sequences comprising the oil body targeting sequence and any one of a variety of proteins or peptides of interest to be expressed and targeted to oil bodies in seeds.

According to one embodiment of the invention the expression cassette is introduced into a host cell in a form where the expression cassette is stably incorporated into the genome of the host cell. Accordingly it is apparent that one

may also introduce the expression cassette as part of a recombinant DNA sequence capable of replication and or expression in the host cell without the need to become integrated into the host chromosome. Examples of this are found in a variety of vectors such as viral or plasmid vectors capable of replication and expression of proteins in the host cell. One specific example are plasmids that carry an origin of replication that permit high copy number such as the pUC series of *E. coli* plasmids additionally said plasmids modified to contain an inducible promoter such as the *LacZ* promoter inducible by galactose or IPTG.

For embodiments of the invention wherein the production and recovery of the recombinant protein is contemplated from non-plant cells the expression cassette so described above is modified to comprise a first recombinant DNA sequence comprising a transcriptional and translational regulatory sequence capable of expression in the intended host production cell or organism. Promoter regions highly active in cells of microorganisms, fungi, insects and animals are well described in the literature of any contemplated host species and may be commercially available or can be obtained by standard methods known to a person skilled in the art. It is apparent that one means to introduce the recombinant molecule to the host cell is through specific infectious entities such as viruses capable of infection of the host modified to contain the recombinant DNA to be expressed.

In a further embodiment of the invention it is contemplated that proteins other than plant oleosins and proteins with homology to plant oleosins that may specifically associate with triglycerides, oils, lipids, fat bodies or any hydrophobic cellular inclusions in the host organism or with reconstituted plant oil bodies may be fused to a recombinant protein and used in the manner contemplated. A system functionally equivalent to plant oleosins and oil bodies has been described in bacteria (Pieper-Fürst et al., 1994, J. Bacteriol. 176:4328 - 4337). Other proteins from additional sources such as, but not limited to; fungi, insects or animals. with equivalent regulatory and targeting properties may be known or discovered by a

person skilled in the art.

Of particular interest for transcriptional and translational regulation in plants of the first recombinant DNA molecule is a regulatory sequence (promoter) from an oil body protein gene, preferably an oil body protein gene expressed in dicotyledonous oil seeds. The expression of these genes in dicotyledonous oilseeds was found to occur much earlier than had hitherto been believed as reported in the literature. Thus, the promoters and upstream elements of these genes are valuable for a variety of uses including the modification of metabolism during phases of embryogenesis which precede the accumulation of storage proteins. Alternatively said promoter may also comprise a promoter capable of expression constitutively throughout the plant or a promoter which has enhanced expression within tissues or organs associated with oil synthesis. Of more particular interest is a promoter that expresses an oil body protein to a high level. Many plant species are tetraploid or hexaploid and may contain numerous copies of functional oil body protein genes. As it is preferable to obtain a gene that is controlled by a promoter that expresses at high levels when compared to other oil body protein genes within the same species it may be advantageous to choose a diploid species as a source of oil body protein genes. An example is the diploid cruciferous plant *Arabidopsis thaliana*, wherein only two or three oil body protein genes are detected by southern blot analysis whereas the seeds contain oil body proteins as a high percentage of total protein.

The degree of evolutionary relationship between the plant species chosen for isolation of a promoter and the plant species selected to carry out the invention may not be critical. The universality of most plant genes and promoter function within dicotyledonous species has been amply demonstrated in the literature. Additionally to a certain extent the conservation of function between monocot and dicot genes has also been shown. This is apparent to a person skilled in the art that the function of any given promoter in any chosen species may be tested prior to practising the invention by simple means such as transient expression of marker gene promoter fusions in isolated cells or intact tissues. The promoter region

typically comprises minimally from 100 bp 5' to the translational start of the structural gene coding sequence, up to 2.5 kb 5' from the same translational start.

Of particular interest as a source of DNA encoding sequences capable of providing for targeting to an oil body protein are oil-body protein genes obtainable from *Arabidopsis* or *Brassica napus* which provide for expression of the protein of interest in seed (See Taylor et al., 1990, Planta 181:18-26). The necessary regions and amino-acid sequences needed to provide targeting to the oil body reside in the highly hydrophobic central region of oil body proteins. The deduced amino acid sequence necessary to provide targeting to the oil body for an *Arabidopsis thaliana* oil-body protein shown in SEQ ID NO.5 is as follows:

10 20
M-M-G-R-D-R-D-Q-Y-Q-M-S-G-R-G-S-D-Y-S-K-
30 40
S-R-Q-I-A-K-A-A-T-A-V-T-A-G-G-S-L-L-V-L-
50 60
S-S-L-T-L-V-G-T-V-I-A-L-T-V-A-T-P-L-L-V-
70 80
I-F-S-P-I-L-V-P-A-L-I-T-V-A-L-L-I-T-G-F-
90 100
L-S-S-G-G-F-G-I-A-A-I-T-V-F-S-W-I-Y-K*Y-
110 120
A-T-G-E-H-P-Q-G-S-D-K-L-D-S-A-R-M-K-L-G-
130 140
S-K-A-Q-D-L-K-D-R-A-Q-Y-Y-G-Q-Q-H-T-G-G-
150
E-H-D-R-D-R-T-R-G-G-Q-H-T-T

Amino acids from about 25-101 comprise the central hydrophobic domain.

To identify other oil body protein genes having the desired characteristics, where an oil body protein has been or is isolated, the protein may be partially

sequenced, so that a probe may be designed for identifying mRNA. Such a probe is particularly valuable if it is designed to target the coding region of the central hydrophobic domain which is highly conserved among diverse species. In consequence, a DNA or RNA probe for this region may be particularly useful for identifying coding sequences of oil body proteins from other plant species. To further enhance the concentration of the mRNA, cDNA may be prepared and the cDNA subtracted with mRNA or cDNA from non-oil body producing cells. The residual cDNA may then be used for probing the genome for complementary sequences, using an appropriate library prepared from plant cells. Sequences which hybridize to the cDNA under stringent conditions may then be isolated.

In some instances, as described above, the use of an oil body protein gene probe (conserved region), may be employed directly for screening a cDNA genomic library and identifying sequences which hybridize to the probe. The isolation may also be performed by a standard immunological screening technique of a seed-specific cDNA expression library. Antibodies may be obtained readily for oil-body proteins using the purification procedure and antibody preparation protocol described by Taylor et al. (1990, Planta, 181:18-26). cDNA expression library screening using antibodies is performed essentially using the techniques of Huynh et al. (1985, in DNA Cloning, Vol. 1, a Practical Approach, ed. D.M. Glover, IRL Press, pp. 49-78). Confirmation of sequence is facilitated by the highly conserved central hydrophobic region (see Figure 1). DNA sequencing by the method of Sanger et al. (1977, Proc. Natl. Acad. Sci. USA, 74:5463-5467) or Maxam and Gilbert (1980, Meth. Enzymol., 65:497-560) may be performed on all putative clones and searches for homology performed. Homology of sequences encoding the central hydrophobic domain is typically 70%, both at the amino-acid and nucleotide level between diverse species. If an antibody is available, confirmation of sequence identity may also be performed by hybrid-select and translation experiments from seed mRNA preparations as described by Sambrook et al. (1990, Molecular Cloning, 2nd Ed., Cold Spring Harbour Press, pp. 8-49 to 8-

51).

cDNA clones made from seed can be screened using cDNA probes made from the conserved coding regions of any available oil body protein gene (e.g., Bowman-Vance and Huang, 1987, J. Biol. Chem., 262:11275-11279). Clones are selected which have more intense hybridization with seed DNAs as compared to seedling cDNAs. The screening is repeated to identify a particular cDNA associated with oil bodies of developing seeds using direct antibody screening or hybrid-select and translation. The mRNA complementary to the specific cDNA is absent in other tissues which are tested. The cDNA is then used for screening a genomic library and a fragment selected which hybridizes to the subject cDNA. Of particular interest for transcriptional and translational regulation in plants of said second recombinant DNA molecule is a regulatory sequence (promoter) from a gene expressed during the germination of seeds and the early stages of growth of a seedling, specifically a gene showing high levels of expression during the stage of mobilization of stored seed reserves, more specifically the promoter sequence from the glyoxisomal enzymes iso-citrate lyase or malate synthase. Information concerning genomic clones of iso-citrate lyase and malate synthase from *Brassica napus* and *Arabidopsis* that have been isolated and described has been published (Comai et al., 1989, Plant Cell 1: 293-300) and can be used by a person skilled in the art, by the methods described above, to isolate a functional promoter fragment. Other enzymes involved in the metabolism of lipids or other seed reserves during germination may also serve as a source of equivalent regulatory regions.

For production of recombinant protein oleosin fusions in heterologous systems such as animal, insect or microbial species, promoters would be chosen for maximal expression in said cells, tissues or organs to be used for recombinant protein production. The invention is contemplated for use in a variety of organisms which can be genetically altered to express foreign proteins including animals, especially those producing milk such as cattle and goats, invertebrates such as insects, specifically insects that can be reared on a large scale, more specifically

those insects which can be infected by recombinant *baculoviruses* that have been engineered to express oleosin fusion proteins, fungal cells such as yeasts and bacterial cells. Promoter regions highly active in viruses, microorganisms, fungi, insects and animals are well described in the literature and may be commercially available or can be obtained by standard methods known to a person skilled in the art. It is preferred that all of the transcriptional and translational functional elements of the initiation control region are derived from or obtained from the same gene.

For those applications where expression of the recombinant protein is derived from extrachromosomal elements, one may chose a replicon capable of maintaining a high copy number to maximize expression. Alternatively or in addition to high copy number replicons, one may further modify the recombinant DNA sequence to contain specific transcriptional or translation enhancement sequences to assure maximal expression of the foreign protein in host cells.

The level of transcription should be sufficient to provide an amount of RNA capable of resulting in a modified seed, cell, tissue, organ or organism. The term "modified " is meant a detectably different phenotype of a seed, cell, tissue, organ or organism in comparison to the equivalent non-transformed material, for example one not having the expression cassette in question in its genome. It is noted that the RNA may also be an "antisense RNA" capable of altering a phenotype by inhibition of the expression of a particular gene.

Ligation of the DNA sequence encoding the targeting sequence to the gene encoding the polypeptide of interest may take place in various ways including terminal fusions, internal fusions, and polymeric and concatameric fusions. In all cases, the fusions are made to avoid disruption of the correct reading frame of the oil-body protein and to avoid inclusion of any translational stop signals in or near the junctions. The different types of terminal an internal fusions are shown in Figure 1 along with a representation of configurations *in vivo*.

In many of the cases described, the ligation of the gene encoding the peptide

preferably would include a linker encoding a protease target motif. This would permit the release of the peptide once extracted as a fusion protein. Potential cleavage sites which could be employed are recognition motifs for thrombin (Leu-Val-Pro-Arg-Gly, SEQ. ID. NO.6) (Fujikawa et al., 1972, Biochemistry 11:4892-4899), of factor Xa (Phe-Glu-Gly-Arg-aa, SEQ. ID NO.7) (Nagai et al., 1985, Proc. Natl Acad. Sci. USA, 82:7252-7255) or collagenase (Pro-Leu-Gly-Pro, SEQ. ID. NO.8) (Scholtissek and Grosse, 1988, Gene 62:55-64). Additionally, for uses where the fusion protein contains a peptide hormone that is released upon ingestion, the protease recognition motifs may be chosen to reflect the specificity of gut proteases to simplify the release of the peptide.

For those uses where chemical cleavage of the polypeptide from the oil body protein fusion is to be employed, one may alter the amino acid sequence of the oil body protein to include or eliminate potential chemical cleavage sites. For example, one may eliminate the internal methionine residues in the *Arabidopsis* oleosin at positions 11 and 117 by site directed mutagenesis to construct a gene that encodes a oleosin that lacks internal methionine residues. By making a N-terminal fusion with the modified oleosin via the N-terminal methionine residue already present in the *Arabidopsis* oleosin, one may cleave the polypeptide of interest by the use of cyanogen bromide providing there are no internal methionines in said polypeptide. Similar strategies for other chemical cleavage agents may be employed. It should be noted that a variety of strategies for cleavage may be employed including a combination of chemical modification and enzymatic cleavage.

By appropriate manipulations, such as restriction, chewing back or filling in overhangs to provide blunt ends, ligation of linkers, or the like, complementary ends of the fragments can be provided for joining and ligation. In carrying out the various steps, cloning is employed, so as to amplify the amount of DNA and to allow for analyzing the DNA to ensure that the operations have occurred in proper manner. A wide variety of cloning vectors are available, where the cloning vector

includes a replication system functional in *E. coli* and a marker which allows for selection of the transformed cells. Illustrative vectors include pBR332, pUC series, M13mp series, pACYC184, etc for manipulation of the primary DNA constructs. Thus, the sequence may be inserted into the vector at an appropriate restriction site(s), the resulting plasmid used to transform the *E. coli* host, the *E. coli* grown in an appropriate nutrient medium and the cells harvested and lysed and the plasmid recovered. Analysis may involve sequence analysis, restriction analysis, electrophoresis, or the like. After each manipulation the DNA sequence to be used in the final construct may be restricted and joined to the next sequence, where each of the partial constructs may be cloned in the same or different plasmids.

The mode by which the oil body protein and the protein to be expressed are fused can be either a N-terminal, C-terminal or internal fusion. The choice is dependant upon the application. For example, C-terminal fusions can be made as follows: A genomic clone of an oil body protein gene preferably containing at least 100 bp 5'to the translational start is cloned into a plasmid vehicle capable of replication in a suitable bacterial host (e.g., pUC or pBR322 in *E. coli*). A restriction site is located in the region encoding the hydrophilic C-terminal portion of gene. In a plant oil body protein of approximately 18 KDa, such as the *Arabidopsis* oleosin, this region stretches typically from codons 125 to the end of the clone. The ideal restriction site is unique, but this is not absolutely essential. If no convenient restriction site is located in this region, one may be introduced by site-directed mutagenesis. The only major restriction on the introduction of this site is that it must be placed 5' to the translational stop signal of the OBP clone.

With this altered clone in place, a synthetic oligonucleotide adapter may be produced which contains coding sequence for a protease recognition site such as Pro-Leu-Gly-Pro or a multimer thereof. This is the recognition site for the protease collagenase. The adaptor would be synthesized in such a way as to provide a 4-base overhang at the 5' end compatible with the restriction site at the 3' end of the oil body protein clone, a 4-base overhang at the 3' end of the adaptor to facilitate

ligation to the foreign peptide coding sequence and additional bases, if needed, to ensure no frame shifts in the transition between the oil body protein coding sequence, the protease recognition site and the foreign peptide coding sequence. The final ligation product will contain an almost complete oil body protein gene, coding sequence for collagenase recognition motif and the desired polypeptide coding region all in a single reading frame.

A similar approach is used for N-terminal fusions. The hydrophilic N-terminal end of oil-body proteins permits the fusion of peptides to the N-terminal while still assuring that the foreign peptide would be retained on the outer surface of the oil body. This configuration can be constructed from similar starting materials as used for C-terminal fusions, but requires the identification of a convenient restriction site close to the translational start of the oil body protein gene. A convenient site may be created in many plant oil body protein genes without any alteration in coding sequence by the introduction of a single base change just 5' to the start codon (ATG). In plant oil body proteins thus far studied, the second amino acid is alanine whose codon begins with a "G". A-C transition at that particular "G" yields a *Nco I* site. As an illustration of such a modification, the context of the sequences is shown below:

3' . .TC TCA ACA ATG GCA . . . Carrot Oil Body Protein (SEQ. ID. NO.9)

3' . .CG GCA GCA ATG GCG . . . Maize 18KDa Oil Body Protein (SEQ. ID. NO.10)

A single base change at the adenine prior to the 'ATG' would yield in both cases CCATGG which is an *Nco I* site. Thus, modification of this base using the site-directed mutagenesis will introduce a *Nco I* site which can be used directly for the insertion of a DNA coding sequence assuming no other *Nco I* sites are present in the sequence. Alternatively other restriction sites may be used or introduced to obtain cassette vectors that provide a convenient means to introduce foreign DNA.

The coding sequence for the foreign peptide may require preparation which will allow its ligation directly into the introduced restriction site. For example,

introduction of a coding sequence into the *Nco I* site introduced into the oil body protein coding sequences described above may require the generation of compatible ends . This may typically require a single or two-base modification by site-directed mutagenesis to generate an *Nco I* site around the translational start of the foreign peptide. This peptide is then excised from its cloning vehicle using *Nco I* and a second enzyme which cuts close to the translational stop of the target. Again, using the methods described above, a second convenient site can be introduced by site-directed mutagenesis. It has been suggested by Qu and Huang (1990, supra) that the N-terminal methionine might be removed during processing of the plant oil body proteins protein *in vivo* and that the alanine immediately downstream of this might be acylated. To account for this possibility, it may be necessary to retain the Met-Ala sequence at the N-terminal end of the protein. This is easily accomplished using a variety of strategies which introduce a convenient restriction site into the coding sequence in or after the Ala codon.

The resultant constructs from these N-terminal fusions would contain an oil body protein promoter sequence, an in-frame fusion in the first few codons of the oil body protein gene of a high value peptide coding sequence with its own ATG as start signal if necessary and the remainder of the oil body protein gene and terminator.

A third type of fusion involves the placing of a high value peptide coding sequence internally to the coding sequence of the oil body protein. This type of fusion requires the same strategy as in N-terminal fusions, but may only be functional with modifications in regions of low conservation, as it is believed that regions of high conservation in these oil body proteins are essential for targeting of the mature protein. A primary difference in this kind of fusion is the necessity for flanking protease recognition sites for the release of the protein. This means that in place of the single protease recognition site thus far described, it is necessary to have the protein of interest flanked by one or more copies of the protease recognition site.

Various strategies are dependant on the particular use and DNA sequence of the inserted coding region and would be apparent to those skilled in the art. The preferred method would be to use synthetic oligonucleotides as linkers to introduce the high value peptide coding sequence flanked by appropriate restriction sites or linkers. Orientation is checked by the use of an asymmetrically placed restriction site in the high-value peptide coding sequence.

The recombinant polypeptide of interest to be produced as an oleosin fusion by any of the specific methods described herein, may be any peptide or protein. For example, proteins that alter the amino acid content of seeds may be used. These include genes encoding proteins high in essential amino acids or amino acids that are limiting in diets, especially arginine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine. Storage proteins such as the high lysine 10 KDa zein from *Zea mays* or the 2S high methionine Brazil Nut storage protein may be used. Alternatively synthetic or modified storage proteins may be employed such as peptides encoding poly-lysine or poly-phenylalanine or fusions of one or more coding regions high in essential amino acids. Proteins may also encode useful additives for animal feeds. These proteins may be enzymes for modification of phytate content in meal such as phytase, more specifically phytase from novel sources and having novel activities. Proteins may also encode hormones useful for boosting productivity such as growth hormones or bovine somatotropin. Proteins may also encode peptides useful for aquaculture.

Proteins may also be those used for various industrial processes. Examples of such proteins include chitinase, glucose isomerase, collagenase, amylase, xylanase, cellulase, lipase, chymosin, renin or various proteases or protease inhibitors. One may also express proteins of interest to the cosmetic industry such as collagen, keratin or various other proteins for use in formulation of cosmetics. Proteins of use to the food industry may also be synthesized including sweetener proteins such as thaumatin, and other flavour enhancing proteins. Proteins that have adhesive properties may also be used.

Of particular interest are those proteins or peptides that may have a therapeutic or diagnostic value. These proteins include antigens, such as viral coat proteins or microbial cell wall or toxin proteins or various other antigenic peptides, peptides of direct therapeutic value such as interleukin-1-ß, the anticoagulant hirudin, blood clotting factors and bactericidal peptides, antibodies, specifically a single-chain antibody comprising a translational fusion of the VH or VL chains of an immunoglobulin. Human growth hormone may also be produced. The invention is not limited by the source or the use of the recombinant polypeptide.

The DNA sequence encoding the polypeptide of interest may be synthetic, naturally derived, or a combination thereof. Dependent upon the nature or source of the DNA encoding the polypeptide of interest, it may be desirable to synthesize the DNA sequence with codons that represent the preference of the organism in which expression takes place. For expression in plant species, one may employ plant preferred codons. The plant preferred codons may be determined from the codons of highest frequency in the proteins expressed in the largest amount in the particular plant species of interest as a host plant.

The termination region which is employed will be primarily one of convenience, since in many cases termination regions appear to be relatively interchangeable. The termination region may be native to the transcriptional initiation region, may be native to the DNA sequence encoding the polypeptide of interest, or may be derived from another source. Convenient termination regions for plant cell expression are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. Termination signals for expression in other organisms are well known in the literature.

A variety of techniques are available for the introduction of DNA into host cells. For example, the chimeric DNA constructs may be introduced into host cells obtained from dicotyledonous plants, such as tobacco, and oleaginous species, such as *Brassica napus* using standard *Agrobacterium* vectors by a transformation protocol such as that described by Moloney et al., 1989, Plant Cell Rep., 8:238-242

or Hinchee et al., 1988, Bio/Technol., 6:915-922; or other techniques known to those skilled in the art. For example, the use of T-DNA for transformation of plant cells has received extensive study and is amply described in EPA Serial No. 120,516; Hoekema et al., 1985. Chapter V, In: The Binary Plant Vector System Offset-drukkerij Kanters B.V., Alblasserdam; Knauf, et al., 1983, Genetic Analysis of Host Range Expression by *Agrobacterium*, p. 245, In: Molecular Genetics of the Bacteria-Plant Interaction, Puhler, A. ed., Springer-Verlag, NY; and An et al., 1985, EMBO J., 4:277-284. Conveniently, explants may be cultivated with *A. tumefaciens* or *A. rhizogenes* to allow for transfer of the transcription construct to the plant cells. Following transformation using *Agrobacterium* the plant cells are dispersed in an appropriate medium for selection, subsequently callus, shoots and eventually plantlets are recovered. The *Agrobacterium* host will harbour a plasmid comprising the *vir* genes necessary for transfer of the T-DNA to the plant cells. For injection and electroporation, (see below) disarmed Ti-plasmids (lacking the tumour genes, particularly the T-DNA region) may be introduced into the plant cell.

The use of non-*Agrobacterium* techniques permits the use of the constructs described herein to obtain transformation and expression in a wide variety of monocotyledonous and dicotyledonous plants and other organisms. These techniques are especially useful for species that are intractable in an *Agrobacterium* transformation system. Other techniques for gene transfer include biolistics (Sanford, 1988, Trends in Biotech., 6:299-302), electroporation (Fromm et al., 1985, Proc. Natl. Acad. Sci. USA, 82:5824-5828; Riggs and Bates, 1986, Proc. Natl. Acad. Sci. USA 83 5602-5606 or PEG-mediated DNA uptake (Potrykus et al., 1985, Mol. Gen. Genet., 199:169-177).

In a specific application, such as to *Brassica napus*, the host cells targeted to receive recombinant DNA constructs typically will be derived from cotyledonary petioles as described by Moloney et al., 1989, Plant Cell Rep., 8:238-242). Other examples using commercial oil seeds include cotyledon transformation in soybean explants (Hinchee et al., 1988, Bio/technology, 6:915-922) and stem transformation

of cotton (Umbeck et al., 1981, Bio/technology, 5:263-266).

Following transformation, the cells, for example as leaf discs, are grown in selective medium. Once shoots begin to emerge, they are excised and placed onto rooting medium. After sufficient roots have formed, the plants are transferred to soil. Putative transformed plants are then tested for presence of a marker. Southern blotting is performed on genomic DNA using an appropriate probe, for example an *A. thaliana* oleosin gene, to show that integration of the desired sequences into the host cell genome has occurred.

The expression cassette will normally be joined to a marker for selection in plant cells. Conveniently, the marker may be resistance to a herbicide, eg phosphinthricin or glyphosate, or more particularly an antibiotic, such as kanamycin, G418, bleomycin, hygromycin, chloramphenicol, or the like. The particular marker employed will be one which will allow for selection of transformed cells compared with cells lacking the introduced recombinant DNA.

The fusion peptide in the expression cassette constructed as described above, expresses at least preferentially in developing seeds. Accordingly, transformed plants grown in accordance with conventional ways, are allowed to set seed. See, for example, McCormick et al. (1986, Plant Cell Reports, 5:81-84). Northern blotting can be carried out using an appropriate gene probe with RNA isolated from tissue in which transcription is expected to occur such as a seed embryo. The size of the transcripts can then be compared with the predicted size for the fusion protein transcript.

Oil-body proteins are then isolated from the seed and analyses performed to determine that the fusion peptide has been expressed. Analyses can be for example by SDS-PAGE. The fusion peptide can be detected using an antibody to the oleosin portion of the fusion peptide. The size of the fusion peptide obtained can then be compared with predicted size of the fusion protein.

Two or more generations of transgenic plants may be grown and either crossed or selfed to allow identification of plants and strains with desired

phenotypic characteristics including production of recombinant proteins. It may be desirable to ensure homozygosity of the plants, strains or lines producing recombinant proteins to assure continued inheritance of the recombinant trait. Methods of selecting homozygous plants are well know to those skilled in the art of plant breeding and include recurrent selfing and selection and anther and microspore culture. Homozygous plants may also be obtained by transformation of haploid cells or tissues followed by regeneration of haploid plantlets subsequently converted to diploid plants by any number of known means, (eg: treatment with colchicine or other microtubule disrupting agents).

The desired protein can be extracted from seed that is preferably homozygous for the introduced trait by a variety of techniques, including use of an aqueous, buffered extraction medium and a means of grinding, breaking, pulverizing or otherwise disrupting the cells of the seeds. The extracted seeds can then be separated (for example, by centrifugation or sedimentation of the brei) into three fractions: a sediment or insoluble pellet, an aqueous supernatant, and a buoyant layer comprising seed storage lipid and oil bodies. These oil bodies contain both native oil body proteins and chimeric oil body proteins, the latter containing the foreign peptide. The oil bodies are separated from the water-soluble proteins and re-suspended in aqueous buffer.

If a linker comprising a protease recognition motif has been included in the expression cassette, a protease specific for the recognition motif is added to the resuspension buffer. This releases the required peptide into the aqueous phase. A second centrifugation step will now re-float the processed oil bodies with their attached proteins and leave an aqueous solution of the released peptide or protein. The foreign protein may also be released from the oil bodies by incubation of the oil body fraction with a different oil body fraction that contains the specific protease fused to oleosin. In this manner the protease cleavage enzyme is removed with the oil bodies that contained the fusion protein with the protease recognition site leaving a product uncontaminated by protease. The desired peptide may be

precipitated, chemically modified or lyophilized according to its properties and desired applications

In certain applications the protein may be capable of undergoing self-release. For example, the proteolytic enzyme chymosin undergoes self-activation from a precursor to an active protease by exposure of the precursor to low pH conditions. Expression of the chymosin precursor/oleosin fusion protein to conditions of low pH will activate the chymosin. If a chymosin recognition site is included between the oleosin and the chymosin protein sequences, the activated chymosin can then cleave the fusion proteins. This is an example of self release that can be controlled by manipulation of the conditions required for enzyme activity. Additional examples may be dependant on the requirement for specific co-factors that can be added when self-cleavage is desired. These may include ions, specific chemical co-factors such as NADH or FADH, ATP or other energy sources, or peptides capable of activation of specific enzymes. In certain applications it may not be necessary to remove the chimeric protein from the oil-body protein. Such an application would include cases where the fusion peptide includes an enzyme which is tolerant to N or C-terminal fusions and retains its activity; such enzymes could be used without further cleavage and purification. The chimeric enzyme/oil body protein would be contacted with substrate as a fusion protein. It is also possible to re-use said oil bodies to process additional substrate as a form of an immobilized enzyme. This specific method finds utility in the batch processing of various substances. The process is also useful for enzymatic detoxification of contaminated water or bodies of water where introduction of freely diffusible enzyme may be undesirable. Said process allows recovery of the enzyme with removal of the oil bodies. It is also possible, if desired, to purify the enzyme - oil body protein fusion protein using an immunoaffinity column comprising an immobilized high titre antibody against the oil body protein.

Other uses for the subject invention are as follows. Oil body proteins comprise a high percentage of total seed protein, thus it is possible to enrich the seed

for certain desirable properties such as high-lysine, high methionine, and the like, simply by making the fusion protein rich in the amino-acid(s) of interest could find utility of particular interest is the modification of grains and cereals which are used as either directly or indirectly as food sources for livestock, including cattle, poultry, and humans. It may be possible to include, as the fusion peptide, an enzyme which may assist in subsequent processing of the oil or meal in conventional oilseed crushing and extraction, for example inclusion of a thermostable lipid-modifying enzyme which would remain active at the elevated crushing temperatures used to process seed and thus add value to the extracted triglyceride or protein product. Other uses of the fusion protein to include use to improve the agronomic health of the crop. For example, an insecticidal protein or a portion of an immunoglobulin specific for an agronomic pest such as a fungal cell wall or membrane, could be coupled to the oil body protein thus reducing attack of the seed by a particular plant pest.

It is possible that the polypeptide/protein will itself be valuable and could be extracted and, if desired, further purified. Alternatively the polypeptide/protein or even the mRNA itself may be used to confer a new biochemical phenotype upon the developing seed. New phenotypes could include such modifications as altered seed-protein or seed oil composition, enhanced production of pre-existing desirable products or properties and the reduction or even suppression of an undesirable gene product using antisense, ribozyme or co-suppression technologies (Izant and Weintraub, 1984, Cell 36: 1007-1015, Hazelhoff and Gerlach, 1988, Nature 334:585-591, Napoli, et al., 1990, Plant Cell, 2:279-289). While one embodiment of the invention contemplates the use of the regulatory sequence in cruciferous plants, it is possible to use the promoter in a wide variety of plant species given the wide conservation of oleosin genes. For example, the promoter could be used in various other dicotyledonous species as well as monocotyledonous plant. A number of studies have shown the spatial and temporal regulation of dicot genes can be conserved when expressed in a monocotyledonous host. The tomato rbcS gene

(Kyozuka et al, 1993, Plant Physiol. 102:991-1000) and the Pin2 gene of potato (Xu et al, 1993 Plant Physiol. 101:683-687) have been shown to function in a monocotyledonous host consistent with their expression pattern observed in the host from which they were derived. Studies have also indicated expression from some dicotyledonous promoters in monocotyledonous hosts can be enhanced by inclusion of an intron derived from a monocotyledonous gene in the coding region of the introduced gene (Xu et al, 1994, Plant Physiol. 106:459-467). Alternatively, given the wide conservation of oleosin genes, it is possible for the skilled artisan to readily isolate oleosin genes from a variety of host plants according to the methodology described within this specification.

It is expected that the desired proteins would be expressed in all embryonic tissue, although different cellular expression can be detected in different tissues of the embryonic axis and cotyledons. This invention has a variety of uses which include improving the intrinsic value of plant seeds by their accumulation of altered polypeptides or novel recombinant peptides or by the incorporation or elimination of a metabolic step. In its simplest embodiment, use of this invention may result in improved protein quality (for example, increased concentrations of essential or rare amino acids), improved lipid quality by a modification of fatty acid composition, or improved or elevated carbohydrate composition. The invention may also be used to control a seed phenotype such as seed coat color or even the development of seed. In some instances it may be advantageous to express a gene that arrests seed development at a particular stage, leading to the production of "seedless" fruit or seeds which contain large amounts of precursors or mature seed products. Extraction of these precursors may be simplified in this case.

Other uses include the inclusion of fusion proteins that contain antigens or vaccines against disease. This application may be particularly relevant to improvements in health care of fish or other wildlife that is not readily assessable by conventional means as the crushed seed can be converted directly into a convenient food source. Other uses include the addition of phytase to improve the nutritional

properties of seed for monogastric animals through the release of phosphate from stored phytate, the addition of chlorophyllase to reduce undesirable chlorophyll contamination of seed oils, especially canola oil and addition of enzymes to reduce anti-metabolites, pigments or toxins from seeds. Additionally the fusion protein may comprise, an insecticidal or fungicidal protein such as magainin or secropin or a portion of an immunoglobulin specific for an agronomic pest, such as a fungal cell wall or membrane, coupled to the oil body protein thus improving seed resistance to pre and post harvest spoilage.

Applications for the use of chimeric proteins associated with the oil body fraction include as above enzymes that are tolerant of N or C-terminal fusions and retain activity. Enzymes associated with oil body suspensions can be mixed with simple or complex solutions containing enzyme substrates. After conversion of substrates to products the enzyme oleosin fusion is readily recovered by centrifugation and floatation and can be reused an indefinite number of times.

The following examples are offered by way of illustration and not by limitation.

Example 1: Isolation of Plant Oleosin Gene. Oil body proteins can be isolated from a variety of sources. The isolation of a oil body protein gene (oleosin) from the plant species *Arabidopsis thaliana* is described herein. Similar methods may be used by a person skilled in the art to isolate oil body proteins from other sources. In this example, a *Brassica napus* oleosin gene (described by Murphy et al, 1991, Biochim Biophys Acta 1088:86-94) was used to screen a genomic library of *A. thaliana* (cv. Columbia) constructed in the Lamda cloning vector EMBL 3A (Obtained from Stratagene Laboratories) using standard techniques. The screening resulted in the isolation of a EMBL 3A clone (referred to as clone 12.1) containing a 15 kb genomic fragment which contains a oleosin gene from *A. thaliana*. The oleosin gene coding region is contained within a 6.6 kb *Kpn I* restriction fragment of this 15 kb fragment. The 6.6 kb *Kpn I* restriction fragment was further mapped and a 1.8 kb *Nco I / Kpn I* fragment containing the oleosin gene including

approximately 850 nucleotides of 5' sequence, the complete coding sequence and the 3' region was isolated. This 1.8 kb fragment was end filled and subcloned in the *Sma I* site of RFM13mp19. The 1.8 kb insert was further digested with a number of standard restriction enzymes and subcloned in M13mp19 for sequencing. Standard cloning procedures were carried out according to Sambrook et al. (Molecular Cloning: A Laboratory Manual 2nd ed., 1989, Cold Spring Harbour Laboratory Press.) The nucleotide sequence was determined and the 1.8 kb sequence of the *A. thaliana* oleosin gene is presented in Figure 2 and SEQ ID No. 1. This particular DNA sequence codes for a 18 KDa *A. thaliana* oleosin gene. The coding region contains a single intron. This gene was used for the construction of recombinant protein expression vectors. The gene may also be used for screening of genomic libraries of other species.

Example 2: Modification of a Native Oleosin for Expression of Heterologous Proteins. The DNA fragment described in example 1 that contains the oleosin gene and regulatory elements was incorporated into an expression cassette for use with a variety of foreign/alternative genes. The following illustrates the modification made to the native *A. thaliana* oleosin gene, especially the promoter and coding region, in order to use this gene to illustrate the invention. It is contemplated that a variety of techniques can be used to obtain recombinant molecules, accordingly this example is offered by way of illustration and not limitation. The *A. thaliana* oleosin gene described in example 1 was cloned as a 1803 bp fragment flanked by *Nco I* and *Kpn I* sites in a vector called pPAW4. The plasmid pPAW4 is a cloning vehicle derived from the plasmid pPAW1 which is a Bluescript plasmid (Clonetech Laboratories) containing a *Brassica napus* Acetolactate synthase (ALS) gene (Wiersma et al., 1989, Mol Gen Genet. 219:413-420). To construct pPAW4, the plasmid pPAW1 was digested with *Kpn I*. The digested DNA was subjected to agarose gel electrophoresis and the fragment that contained the Bluescript plasmid vector backbone and a 677 base pair portion of the *B. napus* ALS gene was isolated and religated. This plasmid contains the following

unique restriction sites within the insert: *Pst I, Nco I, Hind III* and *Kpn I*. This plasmid was called pPAW4. The 1803 bp *Nco I - Kpn I Arabidopsis* oleosin gene fragment was cloned between the *Nco I* and *Kpn I* sites in pPAW4. The resultant plasmid contained in addition to the Bluescript plasmid sequences, a 142 bp *Pst I - Nco I* fragment derived from the *B. napus* ALS gene and the entire 1803 bp *Arabidopsis* oleosin gene. The 142 bp *Pst I - Nco I* fragment is present only as a "stuffer" fragment as a result of the cloning approach and is not used in oleosin expression constructs.

The resultant plasmid was used to further modify the *Arabidopsis* oleosin gene. Site-directed mutagenesis was used to introduce nucleotide changes at positions -2, -1 and +4 in the DNA sequence shown in figure 2. The changes made were: A to T (nucleotide position -2); A to C (nucleotide position -1) and G to A (nucleotide position +4). These nucleotide changes create a 6 nucleotide *Bsp HI* restriction endonuclease site at nucleotide positions -2 to +4. The *Bsp HI* site (T/CATGA) encompasses the ATG initiation codon and provides a recessed end compatible with *Nco I*. A second modification was made by digestion with the enzymes *Eco RV* and *Msc I* which released a 658 bp fragment containing most of the coding sequence of the native oleosin. This digestion left blunt ends at both the *Eco RV* and *Ms cI* sites. The cut vector was recircularized in the presence of an oligonucleotide linker containing the following unique restriction sites: *Hind III, Bgl II, Sal I, Eco RI* and *Cla I*. The recircularized plasmid containing all the 5' regulatory sequences of the oleosin gene, a transcriptional start site and an initiation codon embedded in a *Bsp HI* site. Thirty-one bases downstream of this is a short polylinker containing unique restriction sites. This plasmid was called pOleoP1. The restriction map of this construct is shown in figure 3.

Introduction of any DNA sequence into pOleoP1, this particular cassette requires that the foreign DNA sequence may have, or be modified to have, a *Bsp HI* or *Nco I* site at the initial ATG position. This will assure conservation of the distance between the "cap" site and the initiator codon. Alternatively restriction site

linkers may be added to facilitate insertion into the cassette. The same restriction site can be chosen for the site of insertion of the 3' end of the gene or linkers may be added to introduce appropriate sites. The complete chimeric construct is then excised using the appropriate restriction enzyme(s) and introduced into an appropriate plant transformation vector.

Example 3: Using the *Arabidopsis* Oleosin Promoter For Controlling Expression in Heterologous Plant Species. To demonstrate expression of the oleosin promoter and to determine the amount of 5' regulatory region required for expression in transgenic plants, a small number of DNA constructs were made that contain the 5' transcriptional initiation region of the *Arabidopsis* oleosin gene joined to the coding region for β-glucuronidase (GUS). These constructs were prepared using PCR. The constructs are designated according to the amount of the oleosin 5' region contained, for example, the 2500 construct has approximately 2500 base pairs of the oleosin 5' region. The constructs were introduced into *Brassica napus* and tobacco and the expression of the β-glucuronidase (GUS) gene was measured as described in detail below. The constructs were made using standard molecular biology techniques, including restriction enzyme digestion, ligation and polymerase chain reaction (PCR). As an illustration of the techniques employed, the construction of the 800 construct is described in detail.

In order to obtain a DNA fragment containing approximately 800 base pairs from the 5' transcriptional initiation region of the *Arabidopsis* oleosin gene in a configuration suitable for ligation to a GUS coding sequence, PCR was used. To perform the necessary PCR amplification, two oligonucleotide primers were synthesized (Milligen-Biosearch, Cyclone DNA synthesizer). The first primer, the 5' primer, was called GVR10 and had the following sequence (also shown in SEQ ID NO.11):

5'-CA<u>CTGCAG</u>*GAACTCTCTGGTAA*-3' (GVR10)

The italicized bases correspond to nucleotide positions -833 to -817 in the sequence reported in Figure 2. The *Pst I* site is underlined. The additional

nucleotides 5' of this sequence in the primer are not identical to the oleosin gene, but were included in order to place a *Pst I* site at the 5' end of the amplification product.

The second primer, the 3' primer, is designated as ALP 1 and has the following sequence (also shown in SEQ ID NO.12):

5'-CTACCCGGGATCC*TGTTTACTAGAGAGAATG*-3' (ALP 1)

This primer contains the precise complement (shown in italics) to the sequence reported in Figure 2 from base -13 to -30. In addition, it contains a further 13 bases at the 5' end added to provide two (overlapping) restriction sites, *Sma I* (recognition CCCGGG) and *BamHI* (recognition GGATCC), at the 3' end of the amplification product to facilitate cloning of the PCR fragment. Both the *Sma I* and *Bam HI* sites are underlined, the *Bam HI* site is delineated by a double underline.

These two primers were used in a PCR amplification reaction to produce DNA fragment containing the sequence between nucleotides -833 and -13 of the oleosin gene that now contains a *Pst I* site at the 5' end and *Sma I* and *Bam HI* sites at the 3' end. The template was the oleosin genomic clone 12.1 described in example 1.

The amplification product was called OLEO p800 and was gel purified and digested with *Pst I*. The digestion product was gel purified and end filled using DNA polymerase Klenow fragment then cut with *Sma I* to produce a blunt ended fragment. This fragment was cloned into the *Sma I* site of pUC19 to yield the plasmid pUC OLEOp800. This plasmid contained the insert oriented such that the end of the amplified fragment which contained the *Pst I* site is proximal to the unique *Hind III* site in the pUC19 cloning vector and the end of the amplified fragment that contains the *Sma I* and *Bam HI* site is proximal to the unique *Eco RI* site in the pUC19. This subclone now contains approximately 800 base pairs of 5' regulatory region from the *Arabidopsis* oleosin gene.

The promoter region contained within the plasmid pUC OLEOp800 was

fused to the reporter gene GUS. This was accomplished by substituting the oleosin promoter region for a heat shock promoter fused to a GUS gene in the plasmid HspGUS1559. HspGUS1559 is a plasmid used as a binary vector in *Agrobacterium*, derived from the vector pCGN 1559 (MacBride and Summerfeldt, 1990, Plant Molecular Biology, 14, 269-276) with an insert containing heat shock promoter (flanked by *Bam HI* sites), the β-glucuronidase open reading frame and a nopaline synthase terminator (derived from pB1221, Jefferson RA in Cloning Vectors 1988, Eds. Pouwels P., Enger-Valk BE, Brammer WJ., Elsevier Science Pub BV, Amsterdam section VII, Ai11). The binary plasmid HspGUS1559 was digested with *Bam HI* which resulted in the release of the heat shock promoter and permitted the insertion of a *Bam HI* fragment in its place. pUC OLEOp800 was then cut with *Bam HI* to yield a promoter fragment flanked by *Bam HI* sites. This fragment was cloned into the *Bam HI* sites of the plasmid HspGUS1559 to yield the *Agrobacterium* binary transformation vector pOLEOp800GUS1559. The other constructs were prepared by the same PCR method described above using the appropriate primers for amplifying the -2500 fragment, the -1200 fragment, the -600 fragment or the -200 fragment. These plasmids was used to transform *Brassica napus* and tobacco. GUS expression assays (Jefferson R.A., 1987, Plant Mol. Biol. Rep. 5 387-405) were performed on the developing seeds and on non-reproductive plant parts as controls. The results in *Brassica napus* expressed as specific activity of GUS enzyme are shown in Table I. The results in tobacco are shown in Table II. GUS expression reported is an average obtained from approximately five seeds from each of approximately five different transgenic plants.

These results demonstrate that the oleosin fragment from -833 to -813 used in the 800 construct contains sufficient information to direct specific expression of a reporter gene in transgenic *Brassica napus* embryos as early as heart stage and that the *Arabidopsis* oleosin promoter is capable of directing transcription in plants other than *Arabidopsis*.

It should be noted that the specific expression demonstrated here does not

depend on interactions with the native terminator of an oleosin gene 3' end. In this example, the 3' oleosin terminator was replaced by a terminator derived from the nopaline synthase gene of *Agrobacterium*. Thus, the sequence in the 800 construct is sufficient to achieve the desired expression profile independent of ancillary sequences.

Example 4: Use of Oleosin Promoter and Coding Sequences to Direct Fusion Proteins to the Oil Body Fraction of Seeds. In this example, we have prepared a transgenic plant which expresses, under the control of the oil body promoter, fusion proteins which associate with oilbodies. The enzymatic properties of the inserted coding sequences are preserved while fused to the oleosin. In this example we use the β-glucuronidase enzyme derived from the microorganism *E. coli.* was fused to the oleosin coding region (referred to as a oleosin/GUS fusion) under the control of the *Arabidopsis* oleosin promoter. In order to create an in-frame GUS fusion with the *Arabidopsis* oleosin, two intermediate plasmids were constructed referred to as pOThromb and pGUSNOS.

The plasmid pOThromb comprises the oleosin 5' regulatory region, the oleosin coding sequence wherein the carboxy terminus of the protein has been modified by addition of a thrombin cleavage site. The plasmid pGUSNOS contains the GUS enzyme coding region followed by the nos terminator polyadenylation signal. These two plasmids were joined to make a fusion protein consisting of the oleosin protein fused to the GUS enzyme by way of a linker peptide that is recognized by the endoprotease thrombin.

These plasmids were constructed using PCR and the specific primers shown below. For the construction of pOThromb, a linker oligonucleotide named GVR01 was synthesized having the DNA sequence (shown in SEQ ID NO.13) of:

```
          10        20        30        40
5'AATCCCATGG ATCCTCGTGG AACGAGAGTA GTGTGCTGGC
CACCACGAGT ACGGTCACGG TC 3'  (GVR01)
          50        60
```

This DNA sequence contains from nucleotides 27-62 sequences complementary to the 3' end of the *Arabidopsis* oleosin coding sequence, from nucleotides 12-26 sequences encoding amino acids that comprise the coding region for a thrombin cleavage site, LVPRGS, and from nucleotides 5-14, the sequence for the restriction sites *Bam HI* and *Nco I*. A second primer referred to as GVR10 was also synthesized and consisting of the following DNA sequence (also shown in SEQ ID NO.11):

10 20

5'-CACTGCAGGAACTCTCTGGTAAGC-3' (GVR10)

This DNA sequence contains from nucleotides 5-24 sequences homologous to the oleosin 5' flanking sequence -834 and -814. These two primers were used to amplify the promoter region (0.8 kb) of the *Arabidopsis* oleosin gene contained in the clone 12.1 described in example 1. The resultant fragment was endfilled and cloned in the *Sma I* site of pUC19. This plasmid was called pOThrom which contained the oleosin promoter region, the oleosin coding sequence followed by a cleavage site for the enzyme thrombin and restriction sites for the insertion of the β-glucuronidase (hereinafter GUS).

In order to create an in frame GUS fusion with the *Arabidopsis* oleosin coding region now contained in pOThrom, a GUS gene with the appropriate restriction site was constructed by the use of PCR. An oligonucleotide referred to as GVR20 was synthesized and containing the following DNA sequence (also shown in SEQ ID NO.14):

10 20

5'-GAGGATCCATGGTACGTCCTGTAGAAACC-3' (GVR20)

This oligonucleotide contains from nucleotides 9-29, sequences complementary to the GUS gene and from nucleotides 3-12 the sequence for the restriction sites *Bam HI* and *Nco I* to facilitate cloning. In order to create these restriction sites the fourth nucleotide of the GUS sequence was changed from T to G changing the TTA codon (Leu) into GTA (Val). The second primer used was the universal sequencing primer comprising the DNA sequence (also shown in SEQ

ID NO.15):

10

5'-GTAAAACGACGGCCAGT-3' (Universal Sequencing Primer)

The GVR20 and the Universal Sequencing Primer were used to amplify the GUS-nopaline synthase terminator region from the plasmid pBI121 (Clontech Laboratories). This fragment was endfilled and cloned in the *Sma I* site of pUC19. This plasmid was called pGUSNOS.

The plasmid pOThromb was digested with *Pst I* and *Nco I*, pGUSNOS was digested with *Nco I* and *Xba I*. The inserts of both these plasmids were ligated simultaneously into pCGN1559 cut with *Xba I* and *Pst I* to generate plasmid pCGOBPGUS. The plasmid pCGOBPGUS contained in the following order, the *Arabidopsis* oleosin 5' regulatory region, the oleosin coding region, a short amino acid sequence at the carboxy end of the oleosin coding sequence comprising a thrombin protease recognition site, the coding region for the β-glucuronidase gene followed by the nos terminator polyadenylation signal. The fusion protein coded for by this particular DNA construct is designated as an oleosin/GUS fusion protein.

This plasmid pCGOBPGUS was digested with *Pst I* and *Kpn I* cloned into the *Pst I* and *Kpn I* sites of pCGN1559 resulting in plasmid pCGOBPGUS which was used as a binary vector in *Agrobacterium* transformation experiments to produce transgenic *B. napus*. Seeds from transgenic *Brassica napus* were obtained and tested for GUS activity. The transformed seeds showed GUS activity specifically associated with the oil body fraction. The results of these experiments are shown in Table III. The data demonstrate specific fractionation of the GUS enzyme to the oil body fraction. This example illustrates the expression and targeting of a bacterial derived enzyme specifically to the oil body fraction of transgenic plants.

One skilled in the art would realize that various modifications can be made to the above method. For example, a constitutive promoter may be used to control

the expression of a oleosin/GUS fusion protein. In particular, the 35S promoter may also be used to control the expression of the oleosin/GUS fusion described above by replacing the *Arabidopsis* oleosin promoter with the 35S promoter from CaMV (available from the vector pBI 221.1, Clonetech Laboratories) in the vector pCGOBPGUS. The resultant vector can contain in the following order, the CaMV 35S promoter, the oleosin coding region, a short amino acid sequence at the carboxy end of the oleosin coding sequence comprising a thrombin protease recognition site, the coding region for the β-glucuronidase gene followed by the nos terminator polyadenylation signal. This plasmid can be inserted into Bin 19 and the resultant plasmid may be introduced into *Agrobacterium*. The resulting strain can be used to transform *B. napus*. GUS activity can be measured in the oil body fraction.

Example 5: Cleavage of Oleosin-Fusion Proteins. In example 4 it was demonstrated that the targeting information contained within the oleosin is sufficient to target the protein oleosin/GUS fusion to the oil body. The oleosin/GUS fusion protein contains an amino acid sequence (LVPRGS), which separates the oleosin from GUS. This sequence is recognized by the protease thrombin, which cleaves this peptide sequence after the arginine (R) amino acid residue. The transgenic seeds containing these oleosin/GUS fusions, were used to demonstrate the general utility of such a method of cleavage of a foreign peptide from intact oil bodies containing oleosin/foreign peptide-fusions. The oil body fraction that contained the oleosin/GUS fusion was resuspended in thrombin cleavage buffer which consisted of 50 mM Tris (pH 8.0), 150 mM NaCl, 2.5 mM $CaCl_2$, 2% Triton X-100 and 0.5 % sarcosyl. Thrombin enzyme was added and the sample was placed for 30 minutes each at 45° C, 50° C and 55° C. Following this incubation oil bodies were recovered and tested for GUS activity. GUS enzymatic activity was found in the aqueous phase following this cleavage and removal of the oil bodies. This is shown in table IV. Western blot analysis confirmed the cleavage of GUS enzyme from the oleosin/GUS fusion protein. This example illustrates the cleavage and

recovery of a active enzyme from a oleosin/enzyme fusion following biosynthesis and recovery of the enzyme in the oil body fraction of transgenic seeds.

Example 6: Use of Fusion Proteins as Reusable Immobilized Enzymes. In this example, oleosin/GUS fusion proteins that were associated with oilbodies were used as immobilized enzymes for bioconversion of substrates. Advantage was taken of the fact that enzymatic properties are preserved while fused to the oleosin and the oleosin is very specifically and strongly associated with the oil bodies even when the oil bodies are extracted from seeds. In this example it is demonstrated that said fusion enzymes can be used repeatedly and recovered easily by their association with the oil bodies. In order to demonstrate the reusable and stable GUS activity of the transgenic seeds, transgenic oil bodies were isolated from mature dry seeds as follows. The *Brassica napus* transgenic seeds containing a oleosin/GUS fusion protein were ground in extraction buffer A which consists of 0.15 M Tricine-KOH pH 7.5, 10 mM KCl, 1 mM $MgCl_2$ and 1 mM EDTA, 4 C to which sucrose to a final concentration of 0.6M was added just before use. The ground seeds in extraction buffer were filtered through four layers of cheesecloth before centrifugation for 10 minutes at 5000 x g at 4 C. The oil bodies present as a surface layer were recovered and resuspended in buffer A containing 0.6M sucrose. This solution was overlaid with an equal volume of Buffer A containing 0.1M sucrose and centrifuged at 18,000 x g for 20 minutes. This procedure was repeated twice with the purified oil body fraction (which contained the oilbodies and oleosin/GUS fusion proteins) and was resuspended in buffer A containing 1mM p-nitrophenyl β-D-glucuronide, a substrate for the GUS enzyme. After incubation, the conversion of the colorless substrate to the yellow p-nitrophenol was used as an indication of GUS activity in the suspensions of transgenic oil bodies. This illustrated the activity of the enzyme is maintained while fused to the oleosin protein and the enzyme is accessible to substrate while attached to the oil bodies. The oil bodies were recovered as described above. No GUS enzyme remained in the aqueous phase after removal of the oil bodies. The oil bodies were then added

to fresh substrate. When the oil bodies were allowed to react with fresh substrate, conversion of substrate was demonstrated. This process was repeated four times with no loss of GUS activity. In parallel quantitative experiments, the amount of methyl umbelliferyl glucuronide (MUG) converted to methyl umbelliferone was determined by fluorimetry, and the oil bodies were recovered by flotation centrifugation and added to a new test tube containing MUG. The remaining buffer was tested for residual GUS activity. This procedure was repeated several times. The GUS enzyme showed 100% activity after using four uses and remained stably associated with the oil body fraction. These results are shown in table V. These experiments illustrate the immobilization and recovery of the active enzyme following substrate conversion. The stability of the GUS activity in partially purified oil bodies was established by measuring the GUS activity of the oil body suspension several weeks in a row. The half-life of the GUS activity when the oil-bodies are stored in extraction buffer at 4°C is more than 3 weeks.

Example 7: Expression of IL-1-β as a Fusion Protein. To further illustrate the utility of the invention, the human protein interleuken 1-b (IL-1-β) was chosen for biosynthesis according the method. IL-1-β consists of 9 amino acids (aa); Val-Gln-Gly-Glu-Glu-Ser-Asn-Asp-Lys (Antoni et al., 1986, J. Immunol. 137:3201-3204 SEQ. ID. NO.16). The strategy for biosynthesis was to place this nine amino acid protein at the carboxy terminus of the native oleosin protein. The strategy further employed the inclusion of a protease recognition site to permit the cleavage of the Il-1-β from the oleosin protein while fused to the oil bodies. In order to accomplish this, a recognition site for the endoprotease Factor Xa was incorporated into the construct. The protease Factor Xa can cleave a protein sequence which contains amino acid sequence ile-glu-gly-arg. Cleavage takes place after the arginine residue. Based on these sequences, an oligonucleotide was synthesized which contained 18 nucleotides of the 3' coding region of the *A. thaliana* oleosin (base position 742-759, coding for the last six amino acids of the native protein), an alanine residue (as a result of replacing the TAA stop codon of the native oleosin

with a GCT codon for alanine), the coding sequence for the Factor Xa cleavage (four codons for the amino acids ile-glu-gly-arg) followed by the coding sequence for IL-1-β. The oligonucleotide further comprised a TAA stop coding after the carboxy terminus lysine residue of IL-1-β and adjacent to this stop codon, a *Sal I* restriction site was added. The IL-1-β coding sequence was designed using optimal codon usage for the *B. napus* and *A. thaliana* oleosin. It is apparent to those skilled in the art that maximal expression is expected when the codon usage of the recombinant protein matches that of other genes expressed in the same plant or plant tissue. This oligonucleotide was inserted into the *Arabidopsis* oleosin gene. The modified oleosin gene was cut with *Pst I* and *Sal I* and joined to the nos terminator to obtain the plasmid called pCGOBPILT. This plasmid contains, in the following order, the *Arabidopsis* oleosin promoter, the oleosin coding sequence, including the intron, and the IL-1-β coding region joined at the carboxy terminus of the oleosin protein through a Factor Xa protease recognition site and the nos terminator polyadenylation signal. This construct was inserted into the binary plasmid Bin 19 (Bevan, M., 1984, Nucl. Acids Res. 12:8711-8721) and the resultant plasmid was introduced into *Agrobacterium*. The resulting strain was used to transform *B. napus* and tobacco plants.

The *Arabidopsis* oleosin/IL-1-β fusion was stably integrated into the genomes of tobacco and *B. napus*. Northern analysis of embryo RNA isolated from different transformed tobacco plants showed the accumulation of *Arabidopsis* oleosin/IL-1-β mRNA.

Oil body proteins from transformed tobacco seeds were prepared, and western blotting was performed. An antibody raised against a 22 KDa oleosin of *B. napus*, was used to detect the *Arabidopsis* oleosin/IL-1-β fusion in the tobacco seeds. This antibody recognizes all the major oleosins in *B. napus* and *A. thaliana*. In addition, this antibody recognizes the tobacco oleosins. In oleosins extracted from transformed tobacco seeds the antibody recognized a 20 KDa-protein, which represents oleosin/IL-1-β fusion oleosin. This fusion protein was not present in the

untransformed tobacco seed. These results demonstrate the accumulation of oleosin/IL-1-β fusion in tobacco. Similar expression and accumulation is seen in *Brassica napus* transformed with the oleosin/IL-1-β fusion gene. These results further exemplify the utility of the method for the expression of heterologous proteins in plants.

Example 8: Expression of Oleosin/Hirudin Gene Fusion in *B. napus*. As a further illustration of the invention, the protein hirudin, derived from the leech (a segmented worm) was synthesized and fused to oleosin. Hirudin is an anti-coagulant which is produced in the salivary glands of the leech *Hirudo medicinalis* (Dodt et al., 1984, FEBS Lett., 65:180-183). The protein is synthesized as a precursor protein (Harvey et al., 1986, Proc. Natl. Acad. Sci. USA 83: 1084-1088) and processed into a 65 amino acid mature protein. The hirudin gene was resynthesized to reflect the codon usage of *Brassica* and *Arabidopsis* oleosin genes and a gene fusion was made with the C-terminal end of the *Arabidopsis* oleosin gene. The gene sequences for oleosin and huridin were separated by codons for an amino acid sequence encoding a Factor Xa endoprotease cleavage site. The resulting plasmid was called pCGOBHIRT. This plasmid contains, in the following order, the promoter region of the *Arabidopsis* oleosin gene, the coding sequence of the oleosin protein including the intron, a factor Xa cleavage site and the resynthesized huridin gene followed by the nos terminator polyadenylation signal. This construct was inserted into the binary plasmid Bin 19 and the resultant plasmid was introduced into *Agrobacterium*. The resulting strain was used to transform *B. napus* and tobacco.

The *Arabidopsis* oleosin/hirudin fusion (OBPHIR) was stably integrated into the genomes of *N. tabacum* and *B. napus* respectively. Northern analysis of embryo RNA isolated from different OBPHIR transformed plants showed the accumulation OBPHIR mRNA in *B. napus* seeds. Monoclonal antibodies raised against hirudin confirmed the stable accumulation of the oleosin/hirudin fusion in the seeds of transformed plants. Transgenic seeds containing an oleosin/hirudin

were assayed after a year of storage at room temperature. No degradation of the oleosin/hirudin protein could be observed demonstrating the stability of the huridin in intact seeds.

The huridin can be cleaved from the oleosin by the use of the Factor Xa cleavage site built into the fusion protein. Upon treatment of the oilbody fraction of transgenic *Brassica napus* seeds, active huridin was released. These results are shown in Table VI. This example illustrates the utility of the invention for the production of heterologous proteins with therapeutic value from non-plant sources.

Example 9: Fusion of Foreign Proteins to the N-terminus of Oleosin In this example, a foreign protein was joined to the oleosin coding region via fusion to the N-terminus of the oleosin. As an illustration of the method, the GUS enzyme was fused in-frame to the *Arabidopsis* oleosin coding region described in example 1. In order to accomplish this, four DNA components were ligated to yield a GUS-oleosin fusion under the control of the oleosin promoter. These were: The oleosin 5' regulatory region, the GUS coding region, the oleosin coding region, and the *nos ter* transcription termination region. These four DNA components were constructed as follows:

The first of these components comprised the oleosin promoter isolated by PCR using primers that introduced convenient restriction sites. The 5' primer was called OleoPromK and comprised the sequence (also shown as SEQ. ID. NO.17):

Nco1
5'-CGC GGT ACC ATGG CTA TAC CCA ACC TCG-3'
Kpn1

This primer creates a convenient Kpn 1 site in the 5' region of the promoter. The 3' primer comprised the sequence (also shown as SEQ. ID. NO.18):

5'-CGC ATCGATGTTCTTGTTTACTAGAGAG-3'
Cla1

This primer creates a convenient *Cla 1* site at the end of the untranslated leader sequence of the oleosin transcribed sequence just prior to the ATG initiation

codon in the native oleosin sequence. These two primers were used to amplify a modified promoter region from the native *Arabidopsis* oleosin gene. Following the reaction, the amplification product was digested with *Kpn 1* and *Cla 1* to yield a 870 bp fragment containing the oleosin promoter and the 5' untranslated leader sequence. This promoter fragment is referred to as Kpn-OleoP-Cla and was ligated in the *Kpn 2-Cla 1* sites of a standard subcloning vector referred to as pBS.

The second DNA component constructed was the GUS coding region modified to introduce the appropriate restriction sites and a Factor Xa cleavage site. In order to accomplish this, the GUS coding region in the vector PBI 221 was used as a template in a PCR reaction using the following primers. The 5' primer was called 5'-GUS-Cla which comprised the following sequence (also shown as SEQ. ID. NO.19):

Nde 1
5'- GCC ATCGATCAT ATG TTA CGT CCT GTA GAA ACC CCA- 3'
Cla 1

The 3' primer was referred to as 3'-GUS-FX-Bam and comprised the following nucleotide sequence (also shown as SEQ. ID. NO.20):

5' CGC GGATCC **TCT TCC TTC GAT TTG TTT GCC TCC CTG** C-3'
Bam Hl Factor Xa
encoding DNA sequence
shown in boldface

This second oligonucleotide also encodes four amino acids specifying the amino acid sequence I-E-G-R, the recognition site for the endoprotease activity of factor Xa. The ampli fication product of approximately 1.8 kb comprises a GUS coding region flanked by a *Cla 1* site at the 5' end and in place of the GUS termination codon, a short nucleotide sequence encoding the four amino acids that comprise the Factor Xa endoprotease activity cleavage site. Following these amino

acid codons is a restriction site for *Bam HI*.

The isolation of the oleosin coding region was also performed using PCR. To isolate this third DNA component, the *Arabidopsis* oleosin genomic clone was used as a template in a reaction that contained the following two primers. The first of these primers is referred to as 5'-Bam-Oleo and has the following sequence (also shown as SEQ. ID. NO.21):

5' CGC GGATCC ATG GCG GAT ACA GCT AGA 3'
Bam HI

The second primer is referred to as 3'-Oleo-Xba and has the following sequence (also shown as SEQ. ID. NO.22):

5' TGC TCT AGA CGA TGA CAT CAG TGG GGT AAC TTA AGT 3'
Xba I

PCR amplification of the genomic clone yielded an oleosin coding region flanked by a *Bam HI* site at the 5' end and a *Xba I* site at the 3' end. This coding sequence was subcloned into the *Bam Hi* and *Xba I* site of the subcloning vector pBS.

The fourth DNA component comprised the nopaline synthetase transcriptional termination region (*nos ter*) isolated from the vector pBI 221 as a blunt-ended *Sst I-EcoRI* fragment cloned into the blunt-ended *Hind III* site of pUC 19. This subclone has a *Xba I* site at the 5' end and a *Hind III* site at the 3' end.

As a first step to assemble these four DNA components, the oleosin coding region and *nos ter* were first jointed by ligation of the *Bam HI-Xba I* fragment of the oleosin coding region with the *Xba I-Hind III* fragment of the *nos ter* into *Bam HI-Hind III* digested pUC 19. This construct yielded a subclone that comprised the oleosin coding region joined to the *nos ter*. As a second step in the assembly of

the DNA components, the oleosin promoter region was then joined to the modified GUS coding region by ligation of the *Kpn 1-Cla 1* oleosin promoter fragment to the *Cla 1-Bam H1* fragment of the GUS coding region modified to contain the Factor Xa recognition site and subcloning these ligated fragments into pUC 19 cut with *Kpn 1* and *Bam H1*.

To assemble all four DNA components, the *Kpn 1-Bam H1* oleosin promoter fused to the GUS coding region was ligated with the *Bam H1-Hind III* oleosin coding region-*nos ter* fragment in a tripartite ligation with *Kpn H1-Hind III* digested *Agrobacterium* binary transformation vector PCGN1559. The resultant transformation vector was called pCGYGON1 and was mobilized into *Agrobacterium tumefaciens* EHA 101 and used to transform *B. napus*. Transformed plants were obtained, transferred to the greenhouses and allowed to set seed. Seeds were analyzed as described by Holbrook et al (1991, Plant Physiology 97:1051-1058) and oil bodies were obtained. Western blotting was used to demonstrate the insertion of the GUS oleosin fusion protein into the oil body membranes. In these experiments, more that 80% of the GUS oleosin fusion protein was associated with the oil body fraction. No degradation of the fusion protein was observed. This example illustrates the utility of the method for the expression and recovery of foreign proteins fused to the N-terminus of oleosin.

ADDITIONAL APPLICATIONS OF THE INVENTION

The above examples describe various proteins that can be fused to oleosin and expressed in oil bodies in the seeds of plants such as *Brassica napus*. The above also provides the methodology to prepare such transgenic plants. Therefore one skilled in the art can readily modify the above in order to prepare fusion proteins containing any desired protein or polypeptide fused to oleosin. Several

examples of other proteins that can be produced according to the present invention are provided below.

a) **Expression of Oleosin/Collagenase Fusion Proteins in *B. napus*.** The bacterial collangenase gene of *Vibrio alginolyticus* (Takeuchi et al., 1992, Biochemical Journal, 281:703-708) may be fused to the carboxy terminus of the *Arabidopsis* oleosin gene. This plasmid may contain, in the following order, the promoter region of the *Arabidopsis* oleosin gene, the coding sequence of the oleosin protein including the intron, a factor Xa cleavage site and the collagenase gene followed by the nos terminator polyadenylation signal. The construct can be inserted into the binary plasmid Bin 19 and the resultant plasmid was introduced into *Agrobacterium*. The resulting strain was used to transform *B. napus* and tobacco. The collagenase enzyme was recovered with the oil body fraction in transgenic seeds.

b) **Production of Oleosin/Xylanase Proteins in *B. napus*.** The xylanase gene of *Trichoderma viride* (Gomes, I., Gomes, J., Steiner, W. and Esterbauer, H., 1992, Applied Microbiology and Biotechnology, 36:5, 701-707) may be fused to the carboxy terminus of the *Arabidopsis* oleosin gene. This plasmid may contain, in the following order, the promoter region of the *Arabidopsis* oleosin gene, the coding sequence of the oleosin protein including the intron, a collagenase cleavage site and the xylanase gene followed by the nos terminator polyadenylation signal. The construct may be inserted into the binary plasmid Bin 19 and the resultant plasmid introduced into *Agrobacterium*. The resulting strain can be used to transform *B. napus*. The xylanase enzyme is recovered with the oil body fraction in transgenic seeds. The xylanase enzyme can be further purified by treatment with collagenase to remove the xylanase enzyme from the oleosin protein.

c) **Combination of Two Oleosin Fusion Proteins to Release a Protein Product from Oil Bodies.** Two different oleosin fusions associated with oil bodies can be used as a means to obtain a final product. For example, a transgenic *B. napus* may be obtained which contains a gene that comprises the GUS enzyme fused to the carboxy terminal of oleosin separated by a collagenase protease recognition site. Oil bodies may be obtained from the seed of this plant. These oil bodies can be mixed with the oil bodies described above, which contains collagenase fused to oleosin. The collagenase activity of the oleosin/collagenase fusion protein oil bodies can release the GUS enzyme from the oleosin/GUS fusion proteins oil bodies. The GUS enzyme remains in the aqueous phase after removal of the oil bodies. No collagenase enzyme or contaminating oleosins will remain associated with the purified GUS enzyme illustrating the utility of the invention in obtaining easily purified proteins.

d) **Expression of a Oleosin/Phytase fusion protein in *B. napus*.** A microbial phytase from a *Aspergillus* may be isolated based on the published sequence (van Gorcom et al, European Patent Application 90202565.9, publication number 0 420 358 A1). This gene can be fused to the carboxy terminus of the oleosin protein using techniques described above and a collagenase recognition protease cleave site may be included to allow for separation of the phytase from the oil body if desired. The construct may contain, in the following order, the promoter region of the *Arabidopsis* oleosin gene, the coding sequence of the oleosin protein including the intron, a collagenase cleavage site and the phytase gene followed by the nos terminator polyadenylation signal. The construct can be inserted into the binary plasmid Bin 19 and the resultant plasmid introduced into *Agrobacterium*. The resulting strain can be used to transform *B. napus*. The seed of the transgenic

plants will contain phytase activity. The phytase activity will be associated with the oil body fraction. The phytase activity is useful for the enhancement of meal for monogastric animal feed. The phytase may be purified by treatment with collagenase as described in a), or the transgenic seed may be used as a feed additive.

e) **Expression of a Oleosin/Chymosin fusion protein.** The enzyme chymosin can be expressed as an oleosin fusion protein by joining the coding sequence for chymosin, (for example, described by Alford et al., 1987, US Patent No. 4,666,847) to the oleosin protein as described above. The construct can be used to transform *B. napus.*

f) **Expression of a Oleosin/Glucose isomerase.** The enzyme glucose isomerase can be expressed as a oleosin fusion protein by joining the coding sequence for the enzyme, (for example, described by Wilhelm Hollenberg, 1985, Nucl. Acid. Res. 13:5717-5722) to the oleosin protein as described above. The construct may be used to transform *B. napus.*

g) **Expression of a Oleosin/Zein Storage Protein Fusion.** In order to provide a more favorable nutritional balance for animal feed, a fusion protein may be constructed between the 10 KDa zein protein (Kirihara et al., 1988, Gene 71: 359-370) from corn which is high in methionine and the oleosin coding region. The fusion construct can be made using standard techniques which join at the C-terminus of the oleosin coding region the codon for amino acid 22 of the coding sequence for the 10 KDa zein. The construct can terminate at codon 151 of the zein sequence. The construct may contain, in the following order, the promoter region of the *Arabidopsis* oleosin gene, the coding sequence of the oleosin protein including the intron, codons 22-151 from the 10 KDa zein gene followed by the

nos terminator polyadenylation signal. The construct may be inserted into the binary plasmid Bin 19 and the resultant plasmid introduced into the *Agrobacterium*. The resulting strain can be used to transform *B. napus*.

h) **Expression of a Oleosin/High Lysine Fusion Protein.** In order to increase the lysine content of transgenic seeds, a polylysine oligonucleotide may be added to the C terminus of the oleosin gene. For example, a repetitive oligonucleotide encoding a polylysine coding sequence can be made by synthesizing a $(AAG)_{20}$ oligonucleotide that is joined to the C terminus of the oleosin gene by replacement of the hirudin coding sequence contained within pCBOGHIRT plasmid described above in example 8 with the polylysine oligonucleotide through the use of cohesive restriction termini. The construct may contain, in the following order, the promoter region of the *Arabidopsis* oleosin gene, the coding sequence of the oleosin protein including the intron, 20 codons for the amino acid lysine followed by the nos terminator polyadenylation signal. The construct may be inserted into the binary plasmid Bin 19 and the resultant plasmid may be introduced into the *Agrobacterium*. The resulting strain can be used to transform *B. napus*.

i) **Expression of an Fungicidal Protein as an Oleosin Fusion Protein.** As a further example of the invention, a oleosin fusion protein may be constructed which encodes a protein that is toxic to fungi. For example, the gene for the enzyme chitinase isolated from tobacco (Melchers et al, 1994, Plant Journal 5:469-480) may be fused to the C-terminus of oleosin under the control of the native oleosin promoter. Included in this construct may be an oligonucleotide that encodes a collagenase recognition site located between the oleosin and chitinase coding regions. The expression of this construct will result in the production of a oleosin/chitinase fusion protein from which the chitinase enzyme can be released

from the oleosin by treatment with collagenase. To this construct may be added a second chimeric gene capable of expression of a collagenase enzyme during seed germination. This second gene can comprise approximately 1.5 Kb of the 5' promoter region for isocitrate lyase, the collagenase coding sequence of *Vibrio alginolyticus* (Takeuchi et al., 1992, Biochemical Journal, 281:703-708) and the nos terminator. Isocitrate lyase is a glyoxysomal enzyme expressed under transcriptional control during early stages of seed germination (Comai et al., 1989, The Plant Cell, 1:293-300). This second construct therefore will express collagenase during the germination of the seed and mobilization of the oil body reserves. Expression of isocitrate lyase is restricted to germination and is not expressed in developing seeds. This second gene, joined to the oleosin/chitinase gene can be inserted into the binary vector Bin 19. The resultant vector may be introduced into *Agrobacterium* and used to transform *Brassica napus* plants. It is noted that the two genes may also be introduced independently or in two different plants which are then combined through sexual crossing. Seed from transgenic plants would be collected and tested for resistance to fungi.

j) Expression of an Oleosin Fusion Protein that Provides Protection from Insect Predation. As a further example of the invention, a fusion oleosin protein may be constructed which encodes a protein toxic to foraging insects. For example, the gene for cowpea trypsin inhibitor (Hilder et al., 1987, Nature, 330:160-163) may be used to replace the chitinase gene described in i). The expression of this construct will result in the production of a oleosin/trypsin inhibitor fusion protein from which the trypsin inhibitor can be released from the oleosin by treatment with collagenase. By replacement of the chitinase gene in i) with the trypsin inhibitor, the construct also contains the collagenase gene under control of the germination

specific promoter from the isocitrate lyase gene. This construct may be inserted into the binary vector Bin 19. The resultant vector can be introduced into *Agrobacterium* and used to transform *Brassica napus* plants. Seed from transgenic plants were collected and tested for resistance to insect predation.

k) **Expression of an Enzyme to Alter Secondary Metabolites in Seeds.** In order to alter specific secondary metabolites in the seed, an enzyme encoding tryptophan decarboxylase (TDC) can be expressed in the seed as a fusion to oleosin. This particular enzyme (DeLuca et al., 1989, Proc. Natl. Acad. Sci. USA, 86:2582-2586), redirects tryptophan into tryptamine and causes a depletion of tryptophan derived glucosinolates. This lowers the amount of the antinutritional glucosinolates in the seed and provides a means to further reduce glucosinolate production in crucifer plant species. To accomplish this, a fusion protein may be constructed between the TDC gene and the oleosin coding region. The construct may contain, in the following order, the promoter region of the *Arabidopsis* oleosin gene, the coding sequence of the oleosin protein including the intron, the TDC gene followed by the nos terminator polyadenylation signal. The construct may be inserted into the binary plasmid Bin 19 and the resultant plasmid introduced into *Agrobacterium*. The resulting strain can be used to transform *B. napus*.

EXPRESSION IN PROKARYOTES

Example 10: Isolation of a *B. napus* Oleosin cDNA The *Arabidopsis* oleosin gene described in Example 1 contains an intron, and as such is not suitable for use in a prokaryotic expression system. In order to express oleosin fusions in a microorganism such as bacteria, a coding sequence devoid of introns must be used. To accomplish this, a *B. napus* cDNA library was made using standard techniques and was used to isolate oleosin cDNAs. Four clones were obtained and were called

pcDNA#7, pcDNA#8, pcDNA#10 and pcDNA#12. These cDNA clones were partly sequenced, and one clone pcDNA#8, was sequenced completely. All the clones showed high levels of identity to oleosins. pcDNA#10 was identical to pcDNA#12, but different from pcDNA#8 and pcDNA#7. The deduced amino acid sequence of the insert of pcDNA#8 is very similar to the *Arabidopsis* oleosin and is shown in figure 4. This coding region of oleosin can be used to isolate other oleosin genes or for expression of oleosin fusions in prokaryotic systems. It also provides a convenient coding region for fusion with various other promoters for heterologous expression of foreign proteins due to the ability of the protein (oleosin) to specifically interact with the oilbody fraction of plant extracts.

Example 11: Expression of a Oleosin/GUS Fusion in the Heterologous Host *E. coli*.

In order to further illustrate the invention, an oleosin/GUS gene fusion was expressed in *E. coli* strain JM109. The oleosin cDNA pcDNA#8 described in example 10 was digested with *Nco I* and ligated into the *Nco I* site of pKKGUS, an expression vector containing the LacZ promoter fused to GUS. The plasmid pKKGUS was constructed by adding the GUS coding region to the vector pKK233 (Pharmacia) to generate the plasmid pKKoleoGUS and the anti-sense construct pKKoeloGUS. This construct is shown in Figure 5. These plasmids were introduced into *E. coli* strain JM109 and expression was induced by IPTG. The *E. coli* cells were prepared for GUS activity measurements. In bacterial cells containing the vector pKKGUS, strong induction of GUS activity is observed following addition of ITPG. In cells containing pKKoleoGUS similar strong induction of GUS activity was seen following addition of IPTG. In cells containing pKKoeloGUS (GUS in the antisense orientation) no induction over background was

observed following the addition of IPTG. These results suggest that the oleosin/GUS fusion is active in bacteria. Although that activity observed for the fusion product is less than the unfused product, the oleosin coding sequence was not optimized for expression in bacteria. It is apparent to those skilled in the art that simple modification of codons or other sequences such as ribosome binding sites could be employed to increase expression. The results are summarized in Table VII.

The fusion protein can be isolated from the bulk of the cellular material by utilizing the ability of the oleosin portion of the fusion proteins to specifically associate with oil bodies.

Table I. Expression of *Arabidopsis* oleosin chimearic promoter constructs in transgenic *Brassica napus*.

Promoter Construct (GUS fusion)	Expression of GUS Activity (pmol/MU/mg protein/min)				
	Early Seed (torpedo)	Root	Leaf	Stem	Late Seed (cotyledon)
2500	7709	444	47	88	11607
1200	1795	-	-	-	8980
800	475	-	-	-	7130
600	144	-	-	-	1365
200	65	260	6	26	11
control	14	300	6	30	14

Oleosin promoter-GUS fusions were constructed as described in example 3. Included are GUS values obtained from a control non-transformed plant. A (-) indicated the tissue was not tested. Units are picomoles of methyl umbelliferone (product) per mg protein per minute.

Table II. Expression of *Arabidopsis* oleosin chimearic promoter constructs in transgenic tobacco (*Nicotiana tabacum*).

Promoter Constructs (GUS fusions)	GUS Activity in Seeds (pmol/MU/mg protein/min)
2500	11330
800	10970
Control	0

Oleosin promoter-GUS fusions were constructed as described in example 3.

Included are GUS values obtained from a control non-transformed plant. Units are picomoles of methyl umbelliferone (product) per mg protein per minute.

Table III. Specific partitioning of GUS/oleosin fusions into oil bodies when expressed in transgenic *Brassica napus* plants.

Plant Number	Percent GUS Activity in Oil Bodies (%)	GUS Activity in Oil bodies	GUS Activity 100,000 X g Supernatant	GUS Activity in 100,000 X g Pellet
A1	88	493	1	67
B7	90	243	5	22
control	0	0	0	0

Plants were transformed with an oleosin/GUS fusion protein under the control of the *Arabidopsis* oleosin promoter. Transformed seeds were obtained and fractionated. The initial fractionation consisted of grinding the seeds in 1.5 mL of buffer A consisting of 15 mM Tricine-KOH, pH 7.5, 10 mM KCl, 1 mM Mg Cl_2, 1 mM EDTA, 100 mM sucrose followed by centrifugation at 14,000 X g for 15 minutes at 4°C. From this three fractions were obtained consisting of a floating oil body layer, an aqueous layer and a pellet. The oil body fraction was recovered and assayed for GUS activity. The remaining aqueous phase was further centrifuged for 2 hours at 100,000 X g. The pellet and supernatant from this centrifugation was also tested for GUS activity. Units are nmol MU per mg protein per min.

Table IV. Cleavage of GUS enzyme from oleosin/GUS fusions associated with oil bodies derived from transgenic *Brassica napus* containing an oleosin/GUS fusion protein.

	GUS Activity (nmol product/mg protein/min)		
Fraction	Before Cleavage	After Cleavage	% Activity
Oil bodies	113	26.4	24
100,000 X g supernatant	14.3	83.6	76
100,000 X g pellet	15.7	-	-

Oil bodies containing an oleosin/GUS fusion protein were subjected to cleavage using the endopeptidase thrombin as described in example 5. Values shown are GUS activities before and after cleavage with thrombin. The values are also expressed as a percentage of total GUS activity released following enzyme fusion. Units are nmol methyl umbelliferone per mg protein/min.

Table V. Reuse of oil body associated enzymatic activities.

# Times Oil Bodies Washed	% GUS Activity	
	Oil bodies	Supernatant
1	100	8 ± 5
2	118 ± 7	5 ± 3
3	115 ± 8	3 ± 4
4	119 ± 8	1 ± 20

Oil bodies containing an oleosin/GUS protein were isolated from the seeds of transgenic *Brassica napus*. The oil bodies were added to the fluorometric GUS substrate MUG and allowed to react for one hour. The oil bodies were then recovered and added to a new tube containing the substrate and allowed to react for

one hour again. This process was repeated a total of four times. The table illustrates the reusable activity of the GUS enzyme while still associated with the oil bodies. Values are normalized to 100% as the GUS activity of original oil body isolates.

Table VI. Recovery of active hirudin following synthesis of hirudin in plant seeds.

Treatment	Thrombin Units Per Assay	Antithrombin Units per mg Oil Body Proteins
Buffer only	0.143	0
Wild-type seed	0.146	0
Wild-type seed + factor Xa	0.140	<0.001
Transformed (uncut)	0.140	<0.001
Transformed + factor Xa	0.0065	0.55

Oil bodies containing a hirudin/GUS fusion protein were isolated according to the method and treated with the endoprotease Factor Xa inhibition assay using N-p tosyl-gly-pro-arg-p-nitro anilide (Sigma). Hirudin activity was measured by the use of a thrombin in the method of Dodt et al (1984, FEBS Lett. 65, 180-183). Hirudin activity is expressed as thrombin units per assay in presence of 255 µg of oil body proteins, and also as antithrombin units per mg oil body protein.

Table VII: Expression of active oleosin/GUS fusions in *E. coli*.

Plasmid	Gus Activity
pKK233-2	2.5
pKKoeloGUS	3.1
pKKoleoGUS	28.1
pkkGUS	118.2

As described in example 22, oleosin/GUS fusions were expressed in *E. coli*. Cells were grown, induced with ITPG and GUS activity measured.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: University Technologies International, Inc.

(ii) TITLE OF INVENTION: Oil-Body Proteins As Carriers Of High-Value Peptides In Plants

(iii) NUMBER OF SEQUENCES: 22

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SMART & BIGGAR
(B) STREET: 439 UNIVERSITY AVENUE, SUITE 2300
(C) CITY: TORONTO
(D) PROVINCE: ONTARIO
(E) COUNTRY: COUNTRY
(F) POSTAL CODE: M5G 1Y8

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: GRAVELLE, MICHELINE
(B) REGISTRATION NUMBER: 40,261

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 416-593-5514
(B) TELEFAX: 416-591-1690

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1800 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: ARABIDOPSIS THALIANA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCATGGCTAT ACCCAACCTC GGTCTTGGTC ACACCAGGAA CTCTCTGGTA AGCTAGCTCC 60

ACTCCCCAGA AACAACCGGC GCCAAATTGC CGGAATTGCT GACCTGAAGA CGGAACATCA 120

TCGTCGGGTC CTTGGGCGAT TGCGGCGGAA GATGGGTCAG CTTGGGCTTG AGGACGAGAC 180

CCGAATCGAG TCTGTTGAAA GGTTGTTCAT TGGGATTTGT ATACGGAGAT TGGTCGTCGA 240

GAGGTTTGAG GGAAAGGACA AATGGGTTTG GCTCTGGAGA AAGAGAGTGC GGCTTTAGAG 300

AGAGAATTGA GAGGTTTAGA GAGAGATGCG GCGGCGATGA CGGGAGGAGA GACGACGAGG 360

ACCTGCATTA TCAAAGCAGT GACGTGGTGA AATTTGGAAC TTTTAAGAGG CAGATAGATT 420

TATTATTTGT ATCCATTTTC TTCATTGTTC TAGAATGTCG CGGAACAAAT TTTAAAACTA 480

AATCCTAAAT TTTTCTAATT TTGTTGCCAA TAGTGGATAT GTGGGCCGTA TAGAAGGAAT 540

CTATTGAAGG CCCAAACCCA TACTGACGAG CCCAAAGGTT CGTTTTGCGT TTTATGTTTC 600

GGTTCGATGC CAACGCCACA TTCTGAGCTA GGCAAAAAAC AAACGTGTCT TTGAATAGAC 660

TCCTCTCGTT AACACATGCA GCGGCTGCAT GGTGACGCCA TTAACACGTG GCCTACAATT 720

GCATGATGTC TCCATTGACA CGTGACTTCT CGTCTCCTTT CTTAATATAT CTAACAAACA 780

CTCCTACCTC TTCCAAAATA TATACACATC TTTTTGATCA ATCTCTCATT CAAAATCTCA 840

TTCTCTCTAG TAAACAAGAA CAAAAAAATG GCGGATACAG CTAGAGGAAC CCATCACGAT 900

ATCATCGGCA GAGACCAGTA CCCGATGATG GGCCGAGACC GAGACCAGTA CCAGATGTCC 960

GGACGAGGAT CTGACTACTC CAAGTCTAGG CAGATTGCTA AAGCTGCAAC TGCTGTCACA 1020

GCTGGTGGTT CCCTCCTTGT TCTCTCCAGC CTTACCCTTG TTGGAACTGT CATAGCTTTG 1080

ACTGTTGCAA CACCTCTGCT CGTTATCTTC AGCCCAATCC TTGTCCCGGC TCTCATCACA 1140

GTTGCACTCC TCATCACCGG TTTTCTTTCC TCTGGAGGGT TTGGCATTGC CGCTATAACC 1200

GTTTTCTCTT GGATTTACAA GTAAGCACAC ATTTATCATC TTACTTCATA ATTTTGTGCA 1260

ATATGTGCAT GCATGTGTTG AGCCAGTAGC TTTGGATCAA TTTTTTTGGT CGAATAACAA 1320

ATGTAACAAT AAGAAATTGC AAATTCTAGG GAACATTTGG TTAACTAAAT ACGAAATTTG 1380

ACCTAGCTAG CTTGAATGTG TCTGTGTATA TCATCTATAT AGGTAAAATG CTTGGTATGA 1440

TACCTATTGA TTGTGAATAG GTACGCAACG GGAGAGCACC CACAGGGATC AGACAAGTTG 1500

GACAGTGCAA GGATGAAGTT GGGAAGCAAA GCTCAGGATC TGAAAGACAG AGCTCAGTAC 1560

TACGGACAGC AACATACTGG TGGGGAACAT GACCGTGACC GTACTCGTGG TGGCCAGCAC 1620

ACTACTTAAG TTACCCCACT GATGTCATCG TCATAGTCCA ATAACTCCAA TGTCGGGGAG 1680

TTAGTTTATG AGGAATAAAG TGTTTAGAAT TTGATCAGGG GGAGATAATA AAAGCCGAGT 1740

TTGAATCTTT TTGTTATAAG TAATGTTTAT GTGTGTTTCT ATATGTTGTC AAATGGTACC 1800

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 173 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: ARABIDOPSIS THALIANA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
1 5 10 15

Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20 25 30

Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35 40 45

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
50 55 60

Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
65 70 75 80

Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
85 90 95

Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
100 105 110

Phe Ser Trp Ile Tyr Lys Tyr Ala Thr Gly Glu His Pro Gln Gly Ser
115 120 125

Asp Lys Leu Asp Ser Ala Arg Met Lys Leu Gly Ser Lys Ala Gln Asp
130 135 140

Leu Lys Asp Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu
145 150 155 160

His Asp Arg Asp Arg Thr Arg Gly Gly Gln His Thr Thr
165 170

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 765 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TGATCGACTC GGTACCCGGG GATCCTCTAG AGTCGCGGAT CCATGGCGGA TACAGCTAGA 60

ACCCATCACG ATGTCACAAG TCGAGATCAG TATCCCCGAG ACCGAGACCA GTATTCTATG 120

ATCGGTCGAG ACCGTGACCA GTACTCTATG ATGGGCCGAG ACCGAGACCA GTACAACATG 180

TATGGTCGAG ACTACTCCAA GTCTAGACAG ATTGCTAAGG CTGTTACCGC AGTCACGGCG 240

GGTGGGTCCC TCCTTGTCCT CTCCAGTCTC ACCCTTGTTG GTACTGTCAT TGCTTTGACT 300

GTTGCCACTC CACTCCTCGT TATCTTTAGC CCAATCCTCG TGCCGGCTCT CATCACCGTA 360

GCACTTCTCA TCACTGGCTT TCTCTCCTCT GGTGGGTTTG CCATTGCAGC TATAACCGTC 420

TTCTCCTGGA TCTATAAGTA CGCAACGGGA GAGCACCCAA TCCTCGTGCC GGCTCTCATC 480

ACCGTAGCAC TTCTCATCAC TGGCTTTCTC TCCTCTGGTG GGTTTGCCAT TGCAGCTATA 540

ACCGTCTTCT CCTGGATCTA TAAGTACGCA ACGGGAGAGC ACCCACAGGG GTCAGATAAG 600

TTGGACAGTG CAAGGATGAA GCTGGGAACC AAAGCTCAGG ATATTAAAGA CAGAGCTCAA 660

TACTACGGAC AGCAACATAC AGGTGGTGAG CATGACCGTG ACCGTACTCG TGGTGGCCAG 720

CACACTACTA TCGAAGGAAG AGCCATGGCG CACCTGCAGG CATGC 765

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 234 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Ser Glu Gln Ile Asp Met Ala Asp Thr Ala Arg Thr His His Asp Val
1 5 10 15

Thr Ser Arg Asp Gln Tyr Pro Arg Asp Arg Asp Gln Tyr Ser Met Ile
20 25 30

Gly Arg Asp Arg Asp Gln Tyr Ser Met Met Gly Arg Asp Arg Asp Gln
35 40 45

Tyr Asn Met Tyr Gly Arg Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys
50 55 60

Ala Val Thr Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser
65 70 75 80

Leu Thr Leu Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu
85 90 95

Leu Val Ile Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala
100 105 110

Leu Leu Ile Thr Gly Phe Leu Ser Ser Gly Gly Phe Ala Ile Ala Ala
115 120 125

Ile Thr Val Phe Ser Trp Ile Tyr Lys Tyr Ala Thr Gly Glu His Pro
130 135 140

Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile Thr Gly Phe
145 150 155 160

Leu Ser Ser Gly Gly Phe Ala Ile Ala Ala Ile Thr Val Phe Ser Trp
165 170 175

Ile Tyr Lys Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys Leu
180 185 190

Asp Ser Ala Arg Met Lys Leu Gly Thr Lys Ala Gln Asp Ile Lys Asp
195 200 205

Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu His Asp Arg
210 215 220

Asp Arg Thr Arg Gly Gly Gln His Thr Thr
225 230

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 154 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly Arg Gly Ser
1 5 10 15

Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr Ala Val Thr
20 25 30

Ala Gly Gly Ser Leu Leu Val Leu Leu Ser Leu Thr Leu Val Gly Thr
35 40 45

Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile Phe Ser Pro
50 55 60

Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile Thr Gly Phe
65 70 75 80

Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val Phe Ser Trp
85 90 95

Ile Tyr Lys Tyr Leu Leu Ile Glu His Pro Gln Gly Ser Asp Lys Leu
100 105 110

Asp Ser Ala Arg Met Lys Leu Gly Ser Lys Ala Gln Asp Leu Lys Asp
115 120 125

Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu His Asp Arg
130 135 140

Asp Arg Thr Arg Gly Gly Gln His Thr Thr
145 150

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: thrombin cleavage

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Leu Val Pro Arg Gly
1 5

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Phe Glu Gly Arg Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: collagenase cleavage

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Pro Leu Gly Pro
1

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: carrot obp

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ACGGTAACAA CTCT
14

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: maize obp

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCGGTAACGA CGGC
14

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CACTGCAGGA ACTCTCTGGT AA
22

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CTACCCGGGA TCCTGTTTAC TAGAGAGAAT G
31

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 62 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AATCCCATGG ATCCTCGTGG AACGAGAGTA GTGTGCTGGC CACCACGAGT ACGGTCACGG 60
TC 62

(2) INFORMATION FOR SEQ ID NO:14:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAGGATCCAT GGTACGTCCT GTAGAAACC
29

(2) INFORMATION FOR SEQ ID NO:15:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTAAAAGCAC GGCCAGT
17

(2) INFORMATION FOR SEQ ID NO:16:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: interleukin-1 beta

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Val Gln Gly Glu Glu Ser Asn Asp Lys
1 5

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CGCGGTACCA TGGCTATACC CAACCTCG
28

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CGCATCGATG TTCTTGTTTA CTAGAGAG
28

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GCCATCGATC ATATGTTACG TCCTGTAGAA ACCCCA
36

(2) INFORMATION FOR SEQ ID NO:20:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGCGGATCCT CTTCCTTCGA TTTGTTTGCC TCCCTGC
37

(2) INFORMATION FOR SEQ ID NO:21:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGCGGATCCA TGGCGGATAC AGCTAGA
27

(2) INFORMATION FOR SEQ ID NO:22:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TGCTCTAGAC GATGACATCA GTGGGGTAAC TTAAGT
36

CLAIMS:

1. A method for the expression of a recombinant polypeptide by a host cell said method comprising:

a) introducing into a host cell a chimeric DNA sequence comprising:

1) a first DNA sequence capable of regulating the transcription in said host cell of

2) a second DNA sequence, wherein said second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to a lipid phase linked in reading frame to (ii) a DNA sequence encoding said recombinant polypeptide; and

3) a third DNA sequence encoding a termination region functional in the host cell; and

b) growing said host cell to produce the recombinant fusion polypeptide.

2. The method according to claim 1 further including separating the recombinant fusion polypeptide from cellular host cell components by selective partitioning into a lipid phase.

3. The method according to claim 1 further including separating the recombinant fusion polypeptide from cellular host components by selective partitioning into a lipid

phase comprising oil bodies.

4. The method according to claim 3 wherein said recombinant fusion polypeptide is separated by addition of oil body components and reconstitution of the oil bodies.

5. The method according to claim 2 further comprising releasing the recombinant polypeptide from the recombinant fusion polypeptide associated with the lipid phase, said method comprising:

c) including in said second DNA sequence (2) between said DNA sequence (i) encoding the oil body protein and the DNA sequence (ii) encoding the recombinant polypeptide, a linker DNA sequence (iii) encoding an amino acid sequence that is specifically cleavable by enzymatic or chemical means; and

d) contacting the lipid phase with said enzymatic or chemical means such that said recombinant polypeptide is released from the recombinant fusion polypeptide.

6. The method according to claim 5 wherein said amino acid sequence encoded by said linker DNA sequence is cleavable by enzymatic means.

7. The method according to claim 6 wherein said linker DNA sequence encodes an amino acid sequence that is recognizable by the proteolytic action of an enzyme selected from the group consisting of thrombin, factor Xa, collagenase and chymosin.

8. The method according to claim 6 wherein said enzymatic means comprises an enzyme that is immobilized.

9. The method according to claim 8 wherein said enzyme is immobilized by attachment to an oil body protein that is associated with an oil body.

10. The method according to claim 1 wherein said recombinant polypeptide is an enzyme.

11. The method according to claim 10 wherein said recombinant polypeptide is an enzyme that retains its enzymatic properties while part of the recombinant fusion polypeptide associated with the oil body.

12. A method for the production and release of a recombinant polypeptide from a recombinant fusion polypeptide associated with a plant oil body fraction during seed germination and plant seedling growth, said method comprising:

a) introducing into a plant cell a first chimeric DNA sequence comprising:

1) a first DNA sequence capable of regulating the transcription in said plant cell of

2) a second DNA sequence wherein said DNA second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion

polypeptide to an oil body, linked in reading frame to (ii) a DNA sequence encoding a recombinant polypeptide and (iii) a linker DNA sequence encoding an amino acid sequence that is specifically cleavable by enzymatic means wherein said linker DNA sequence (iii) is located between said DNA sequence (i) encoding the oil body protein and said DNA sequence (ii) encoding the recombinant polypeptide; and

3) a third DNA sequence encoding a termination region;

b) sequentially or concomitantly introducing into the genome of said plant a second chimeric DNA sequence comprising:

1) a first DNA sequence capable of regulating the transcription specifically during seed germination and seed growth of

2) a second DNA sequence encoding a specific enzyme that is capable of cleaving the linker DNA sequence of said first chimeric DNA sequence; and

3) a third DNA sequence encoding a termination region;

c) regenerating a plant from said plant cell and growing said plant to produce seed whereby said recombinant fusion polypeptide is expressed and associated with oil bodies and

d) allowing said seed to germinate wherein said enzyme in said second chimeric DNA sequence is expressed and cleaves the recombinant polypeptide from the recombinant fusion polypeptide associated with the oil bodies during seed germination and early seedling growth.

13. A method for producing an altered seed meal by producing a recombinant

polypeptide in association with a plant seed oil body fraction. said method comprising:

a) introducing into a plant cell a chimeric DNA sequence comprising:

1) a first DNA sequence capable of regulating the transcription in said plant cell of

2) a second DNA sequence wherein said second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to an oil body, linked in reading frame to (ii) a DNA sequence encoding a recombinant polypeptide and

3) a third DNA sequence encoding a termination region;

b) regenerating a plant from said plant cell and growing said plant to produce seed whereby said recombinant polypeptide is expressed and associated with oil bodies; and

c) crushing said seed and preparing an altered seed meal.

14. A method of preparing an enzyme in a host cell in association with an oil body and releasing said enzyme from the oil body, said method comprising:

a) transforming a host cell with a chimeric DNA sequence comprising:

1) a first DNA sequence capable of regulating the transcription of

2) a second DNA sequence, wherein said second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to an oil body; (ii) a DNA sequence encoding an enzyme and (iii) a linker

DNA sequence located between said DNA sequence (i) encoding the oil body and said DNA sequence (ii) encoding the enzyme and encoding an amino acid sequence that is cleavable by the enzyme encoded by the DNA sequence (ii); and

3) a third DNA sequence encoding a termination region functional in said host cell

b) growing the host cell to produce the recombinant fusion polypeptide under conditions such that enzyme is not active;

c) recovering the oil bodies containing the recombinant fusion polypeptide: and

d) altering the environment of the oil bodies such that the enzyme is activated and cleaves itself from the recombinant fusion polypeptide.

15. The method according to claim 14 wherein said enzyme is activated by lowering the pH of the oil body environment.

16. A method for the expression of a recombinant polypeptide by a host cell in association with an oil body and separating said recombinant polypeptide from the oil body, said method comprising:

a) transforming a first host cell with a first chimeric DNA sequence comprising:

1) a first DNA sequence capable of regulating the transcription in said host cell of

2) a second DNA sequence, wherein said second sequence encodes a first recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient

portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to a lipid phase linked in reading frame to (ii) a DNA sequence encoding said recombinant polypeptide; and (iii) a linker DNA sequence encoding an amino acid sequence that is specifically cleavable by enzymatic means wherein said linker DNA sequence (iii) is located between said (i) DNA sequence encoding the oil body protein and said (ii) DNA sequence encoding the recombinant polypeptide; and

3) a third DNA sequence encoding a termination region functional in the host cell; and

b) transforming a second host cell with a second chimeric DNA sequence comprising:

1) a first DNA sequence capable of regulating the transcription specifically during seed germination and seed growth of

2) a second DNA sequence wherein said second sequence encodes a second recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the second recombinant fusion polypeptide to a lipid phase linked in reading frame to do a DNA sequence. encoding a specific enzyme that is capable of cleaving the linker DNA sequence of said first chimeric DNA sequence; and

3) a third DNA sequence encoding a termination region;

c) growing said first host cell under conditions such that the first recombinant fusion polypeptide is expressed and associated with the oil bodies to produce a first oil body fraction containing the first recombinant fusion polypeptide;

d) growing said second host cell under conditions such that the second recombinant fusion polypeptide is expressed and associated with the oil bodies to product a second oil body fraction containing the second recombinant fusion polypeptide;

e) contacting the first oil body fraction of step (c) with the second oil body fraction of step (d) under conditions such that the enzyme portion of the second recombinant fusion polypeptide cleaves the first recombinant polypeptide from the first recombinant fusion polypeptide.

17. The method according to claim 1 wherein said recombinant polypeptide is an interleukin.

18. The method according to claim 1 wherein said recombinant polypeptide is a thrombin inhibitor.

19. The method according to claim 1 wherein said recombinant polypeptide is hirudin.

20. The method according to claim 1 wherein said host cell is a plant cell.

21. The method according to claim 20 wherein said plant is dicotyledonous.

22. The method according to claims 20 wherein said plant is from the family

Brassicaceae.

23. The method according to claim 1 wherein said host cell is a bacterial cell.

24. The method according to claim 1 wherein said host cell is selected from the group consisting of yeast, fungus, viral, insect and animal cells.

25. The method according to any one of claims 1 to 24 wherein said second DNA sequence (ii) is an oleosin derived from a plant from the family Brassicaceae.

26. The method according to any one of claims 1 to 24 wherein said second DNA sequence (ii) is an oleosin gene derived from *Arabidopsis thaliana.*

27. The method according to claim 26 wherein said first DNA sequence (1) is an oleosin gene derived from *Arabidopsis thaliana.*

28. The method according to claim 26 wherein said DNA sequence (i) has the sequence as shown in SEQ ID NO. 1.

29. The method according to claim 26 wherein said DNA sequence (i) encodes a polypeptide having the amino acid sequence as shown in SEQ ID NO. 5.

30. A chimeric DNA sequence, capable of being expressed in association with an oil body of a host cell comprising:

1) a first DNA sequence capable of regulating the transcription in said host cell of

2) a second DNA sequence, wherein said second sequence encodes a recombinant fusion polypeptide and comprises (i) a DNA sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the recombinant fusion polypeptide to a lipid phase linked in reading frame to (ii) a DNA sequence encoding said recombinant polypeptide; and

3) a third DNA sequence encoding a termination region functional in the host cell.

31. The chimeric DNA sequence according to claim 30 wherein said DNA sequence (ii) encodes an enzyme.

32. The chimeric DNA sequence according to claim 30 further including (iii) a linker DNA sequence encoding an amino acid sequence that is specifically cleavable by enzymatic means wherein said linker DNA sequence (iii) is located between said (i) DNA sequence encoding the oil body protein and said (ii) DNA sequence encoding the recombinant polypeptide.

33. The chimeric DNA according to claim 32 wherein said linker DNA sequence (iii)

encodes a cleavage site for an enzyme selected from the group consisting of thrombin, factor, Xa, collagenase and chymosin.

34. The chimeric DNA according to claim 30 wherein said DNA sequence (ii) encodes an interleukin.

35. The chimeric DNA according to claim 30 wherein said DNA sequence (ii) encodes a thrombin inhibitor.

36. The chimeric DNA according to claim 30 wherein said DNA sequence (ii) encodes hirudin.

37. The chimeric DNA sequence according to claim 30 wherein said DNA sequence (i) is an oleosin derived from a plant from the family *Brassicaceae*.

38. The chimeric DNA according to claim 26 wherein said DNA sequence (i) is oleosin derived from *Arabidopsis thaliana*.

39. The chimeric DNA according to any one of claims 30 to 38 wherein said first DNA sequence (1) is an oleosin gene derived from *Arabidopsis thaliana*.

40. The chimeric DNA according to any one of claims 30 to 38 wherein said DNA

sequence (i) has the sequence as shown in SEQ ID NO. 1.

41. The chimeric DNA according to any one of claims 30 to 38 wherein said DNA sequence (i) encodes a polypeptide having the amino acid sequence as shown in SEQ ID NO. 5.

42. An expression cassette comprising a chimeric DNA sequence according to any one of claims 30 to 38.

43. A plant transformed with a chimeric DNA sequence according to any one of claims 30 to 38.

44. A plant cell culture containing a chimeric DNA sequence according to any one of claims 30 to 38.

45. A plant seed containing a chimeric DNA sequence according to any one of claims 30 to 38.

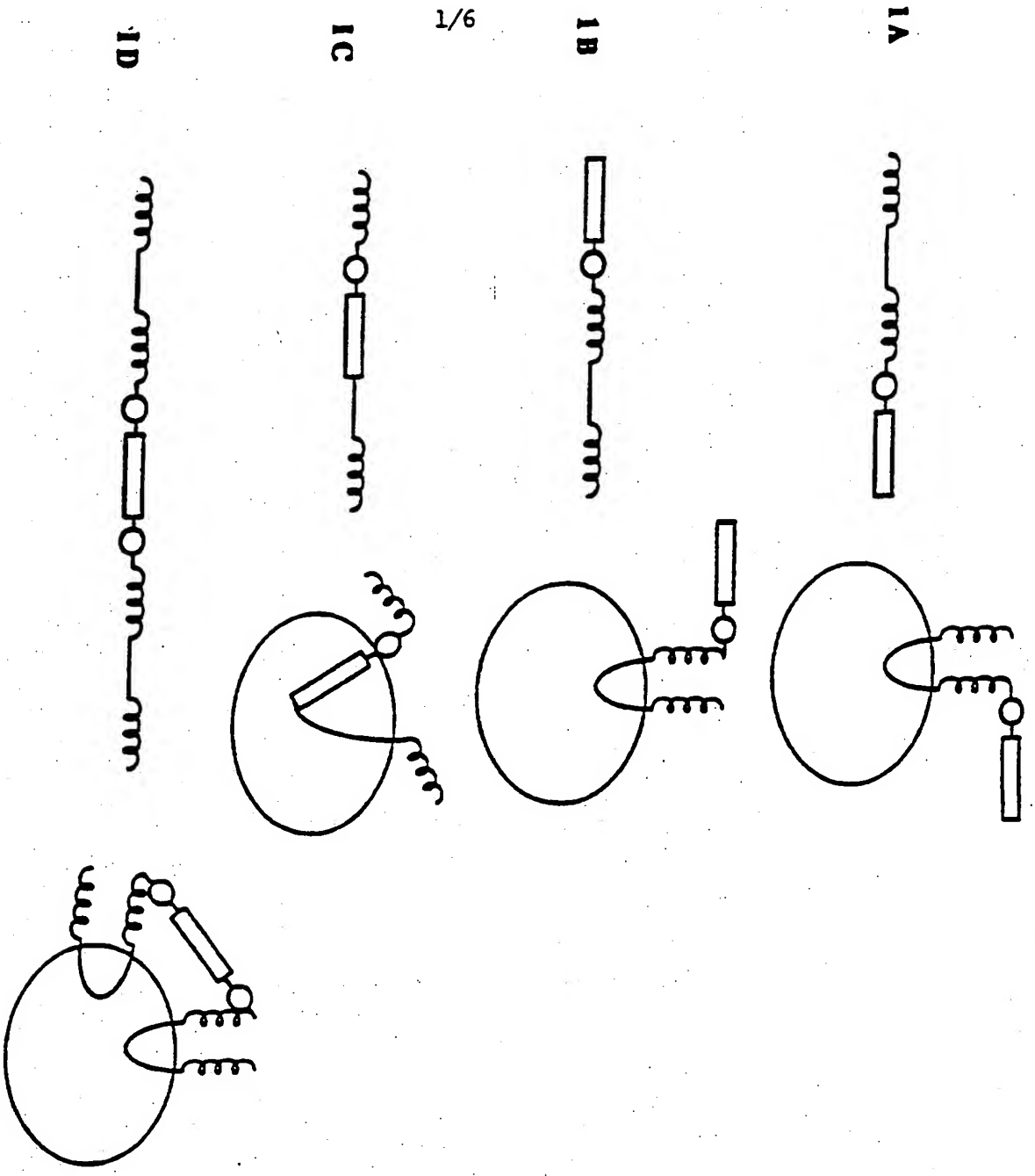


Fig 1

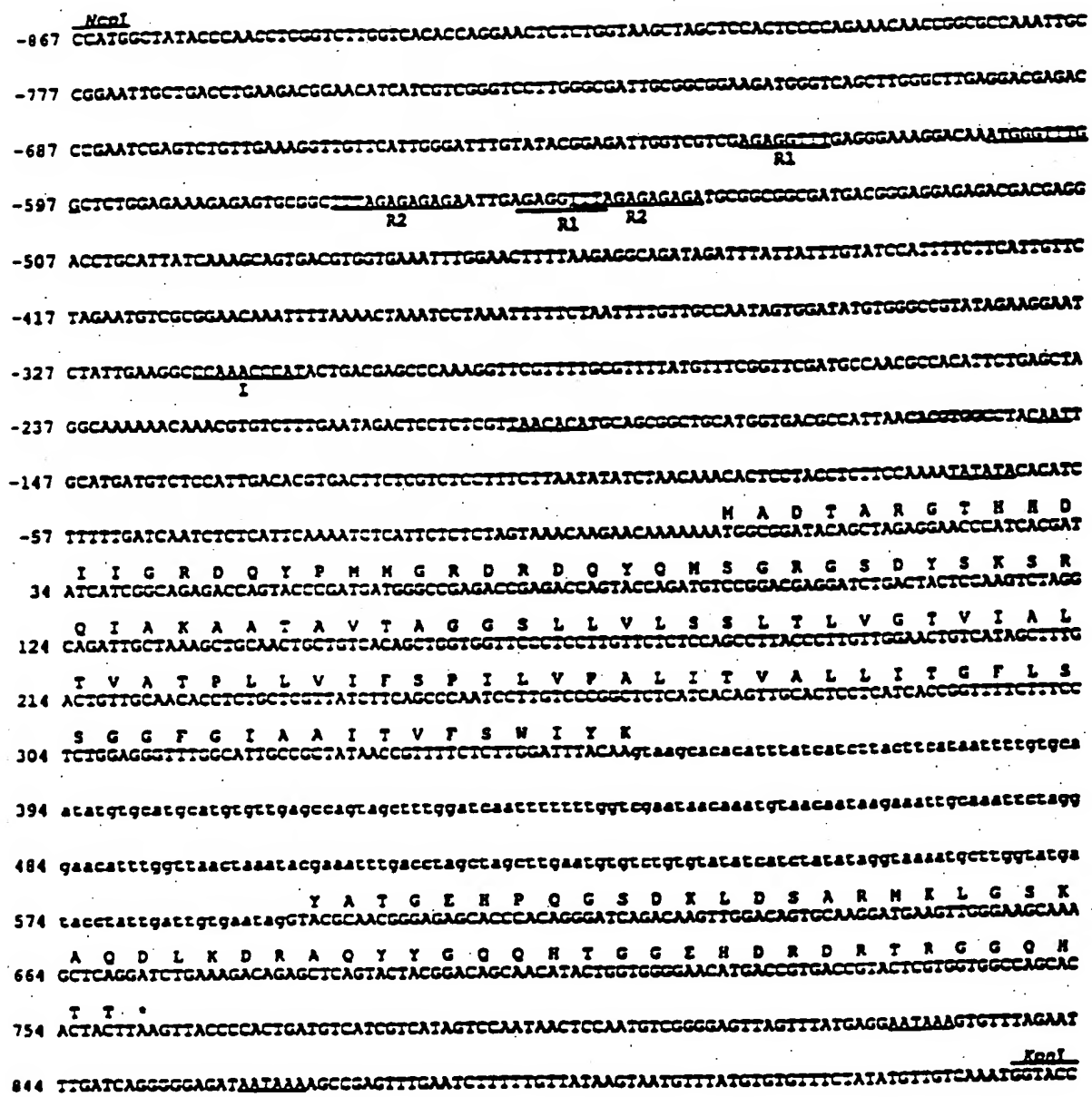

NcoI
R2
R1
R2
R1
I
M A D T A R G T H R D
I I G R D Q Y P M M G R D R D Q Y Q M S G R G S D Y S K S R
Q I A K A A T A V T A G G S L L V L S S L T L V G T V I A L
T V A T P L L V I F S P I L V P A L I T V A L L I T G F L S
S G G F G I A A I T V F S W I Y K
Y A T G E H P Q G S D K L D S A R M K L G S K
A Q D L K D R A Q Y Y G Q Q H T G G E H D R D R T R G G Q H
T T *
KpnI

Fig 2

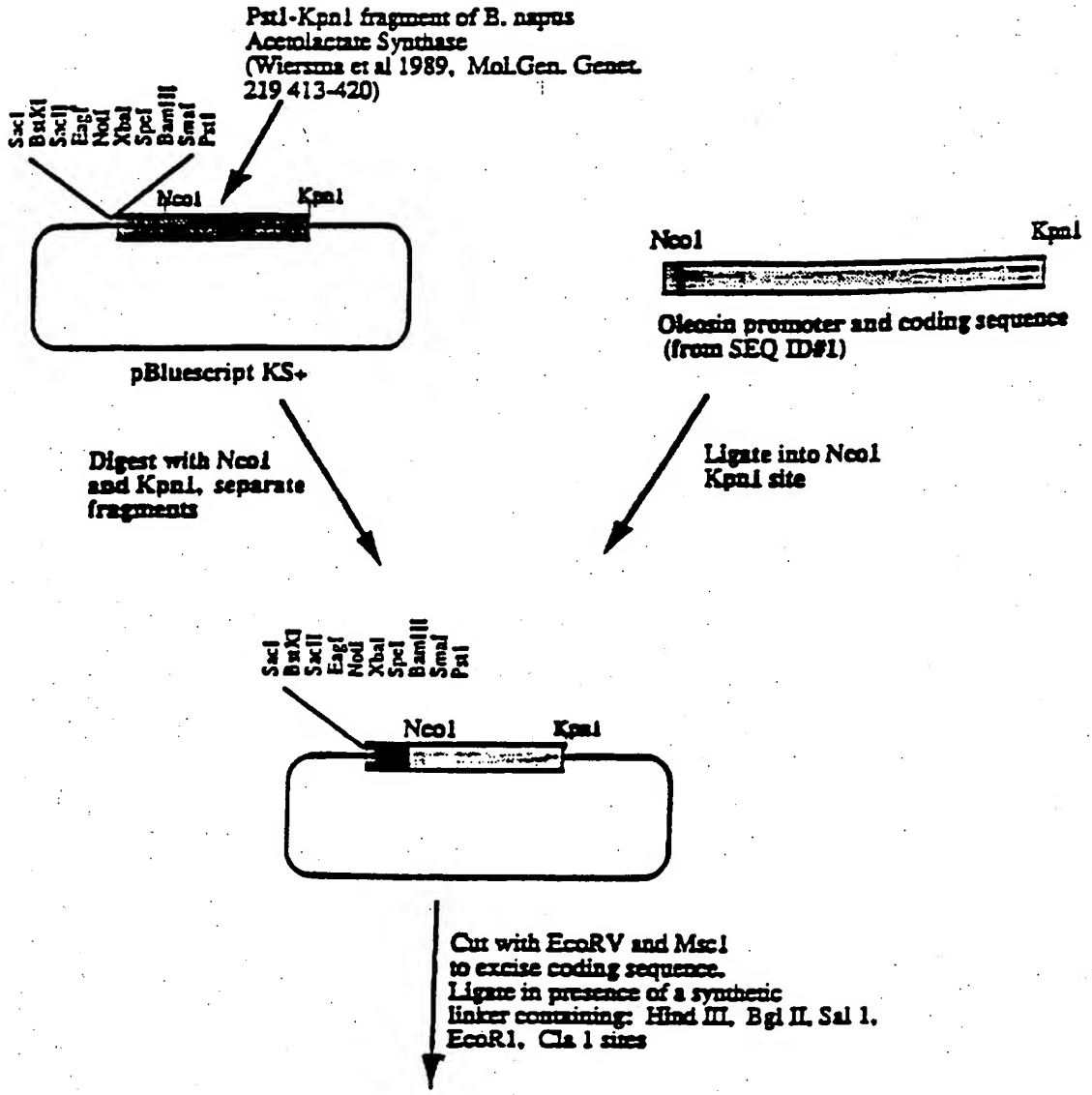
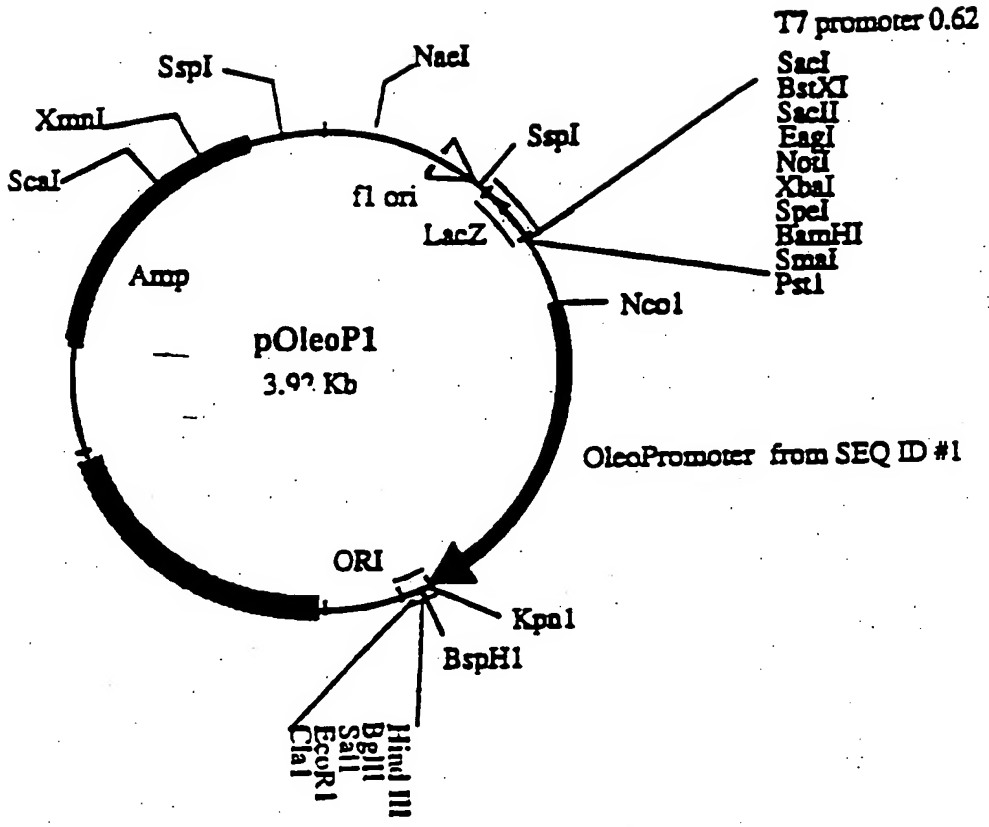

Pst1-Kpn1 fragment of B. napus
Acetolactate Synthase
(Wiersma et al 1989, Mol.Gen. Genet.
219 413-420)
Sac1 BstX1 Sac11 Eag1 Not1 Xba1 Spe1 BamH1 Sma1 Pst1
Nco1
Kpn1
pBluescript KS+
Nco1
Kpn1
Oleosin promoter and coding sequence
(from SEQ ID#1)
Digest with Nco1
and Kpn1, separate
fragments
Ligate into Nco1
Kpn1 site
Sac1 BstX1 Sac11 Eag1 Not1 Xba1 Spe1 BamH1 Sma1 Pst1
Nco1
Kpn1
Cut with EcoRV and Msc1
to excise coding sequence.
Ligate in presence of a synthetic
linker containing: Hind III, Bgl II, Sal 1,
EcoR1, Cla 1 sites

Fig 3 P1



Plasmid name: pOleoP1 (=pBlueScript KS + Oleosin Promoter +Pst1-Nco1 stuffer)

Plasmid size: 3.92 kb
Constructed by: M M Moloney

Fig 3 P2

```
1/1                                                 31/11
ATG GCG GAT ACA GCT AGA ACC CAT CAC GAT GTC ACA AGT CGA GAT CAG TAT CCC CGA GAC
M   A   D   T   A   R   T   H   H   D   V   T   S   R   D   Q   Y   P   R   D
61/21                                               91/31
CGA GAC CAG TAT TCT ATG ATC GGT CGA GAC CGT GAC CAG TAC TCT ATG ATG GGC CGA GAC
R   D   Q   Y   S   M   I   G   R   D   R   D   Q   Y   S   M   M   G   R   D
121/41                                              151/51
CGA GAC CAG TAC AAC ATG TAT GGT CGA GAC TAC TCC AAG TCT AGA CAG ATT GCT AAG GCT
R   D   Q   Y   N   M   Y   G   R   D   Y   S   K   S   R   Q   I   A   K   A
181/61                                              211/71
GTT ACC GCA GTC ACG GGG GGT GGG TCC CTC CTT GTC CTC TCC AGT CTC ACC CTT GTT GGT
V   T   A   V   T   A   G   G   S   L   L   V   L   S   S   L   T   L   V   G
241/81                                              271/91
ACT GTC ATT GCT TTG ACT GTT GCC ACT CCA CTC CTC GTT ATC TTT AGC CCA ATC CTC GTG
T   V   I   A   L   T   V   A   T   P   L   L   V   I   F   S   P   I   L   V
301/101                                             331/111
CCG GCT CTC ATC ACC GTA GCA CTT CTC ATC ACT GGC TTT CTC TCC TCT GGT GGG TTT GCC
P   A   L   I   T   V   A   L   L   I   T   G   F   L   S   S   G   G   F   A
361/121                                             391/131
ATT GCA GCT ATA ACC GTC TTC TCC TGG ATC TAT AAG TAC GCA ACG GCA GAG CAC CCA CAG
I   A   A   I   T   V   F   S   W   I   Y   K   Y   A   T   G   E   H   P   Q
421/141                                             451/151
GGG TCA GAT AAG TTG GAC AGT GCA AGG ATG AAG CTG GGA ACC AAA GCT CAG GAT ATT AAA
G   S   D   K   L   D   S   A   R   M   K   L   G   T   K   A   Q   D   I   K
481/161                                             511/171
GAC AGA GCT CAA TAC TAC GGA CAG CAA CAT ACA GGT GGT GAG CAT GAC CGT GAC CGT ACT
D   R   A   Q   Y   Y   G   Q   Q   H   T   G   G   E   H   D   R   D   R   T
541/181
CGT GGT GGC CAG CAC ACT ACT STOP
R   G   G   Q   H   T   T   *
```

Fig. 4

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/82 C12N15/29 C12N15/57 C12N15/25 C12N15/15
C12N15/62 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,93 21320 (UNIV TECHNOLOGIES INTERNATIONA ;MOLONEY MAURICE M (CA)) 28 October 1993 see page 16, line 4 - line 18 ---	1-8,10, 11, 17-22, 25-45
O,X	TRENDS IN BIOTECHNOLOGY, vol. 13, no. 9, September 1995, pages 379-387, XP002005043 GODDIJN, O.J.M., ET AL.: "Plants as bioreactors" see page 384 right col. section entitled " Reduction of purification costs: oleosin-fusion proteins" and reference 39 & 4TH. INT. CONGR. PLANT MOL. BIOL., HELD JUNE 1994, AMSTERDAM, ABSTRACT 1951., PARMENTER, D.L., ET AL.: --- -/--	19,36

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 June 1996	17. 06. 96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016	Maddox, A

Form PCT/ISA/210 (second sheet) (July 1992)

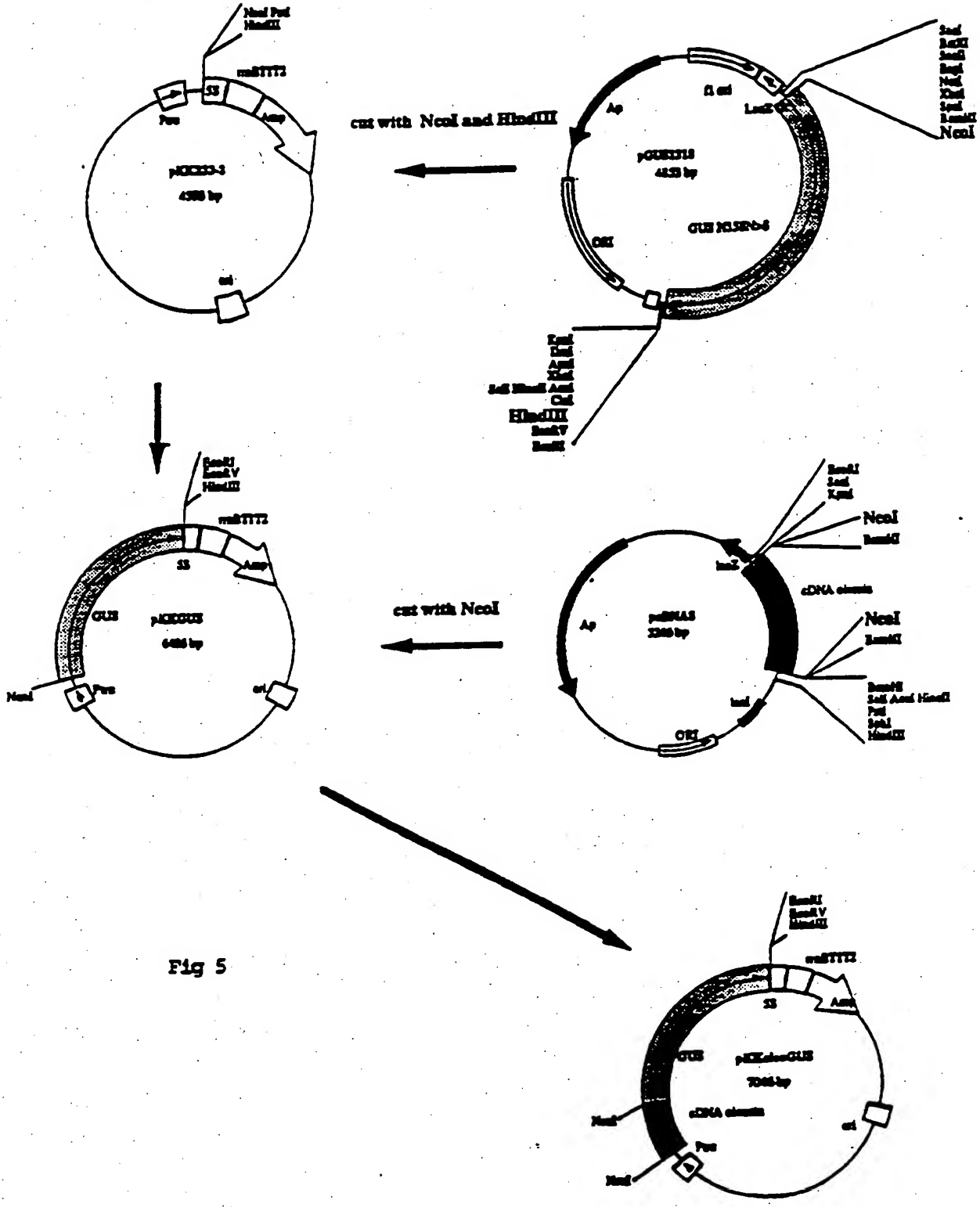


Fig 5

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A-0449376		JP-T-	6501838	03-03-94
		JP-T-	6502296	17-03-94
		WO-A-	9114782	03-10-91
		WO-A-	9114772	03-10-91
US-A-5162601	10-11-92	NONE		

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9321320	28-10-93	AU-B- 1583292	18-11-93
		EP-A- 0636179	01-02-95
		EP-A- 0633948	18-01-95
		FI-A- 944550	30-11-94
		FI-A- 944841	14-12-94
		JP-T- 7507199	10-08-95
		JP-T- 7505525	22-06-95
		NO-A- 943664	15-11-94
		NO-A- 943914	09-12-94
		WO-A- 9320216	14-10-93
WO-A-9320216	14-10-93	WO-A- 9321320	28-10-93
		AU-B- 3884493	08-11-93
		CA-A- 2133417	14-10-93
		EP-A- 0633940	18-01-95
		FI-A- 944550	30-11-94
		JP-T- 7505525	22-06-95
		NO-A- 943664	15-11-94
EP-A-0348348	27-12-89	AU-B- 631551	03-12-92
		AU-B- 3656889	21-12-89
		JP-A- 2046238	15-02-90
EP-A-0449375	02-10-91	AU-B- 649447	26-05-94
		AU-B- 7765691	21-10-91
		AU-B- 632941	14-01-93
		AU-B- 7776691	21-10-91
		CA-A- 2056396	24-09-91
		EP-A- 0449376	02-10-91
		JP-T- 6501838	03-03-94
		JP-T- 6502296	17-03-94
		WO-A- 9114782	03-10-91
		WO-A- 9114772	03-10-91
EP-A-0449376	02-10-91	AU-B- 649447	26-05-94
		AU-B- 7765691	21-10-91
		AU-B- 632941	14-01-93
		AU-B- 7776691	21-10-91
		CA-A- 2056396	24-09-91
		EP-A- 0449375	02-10-91

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO,A,93 20216 (UNIV TECHNOLOGIES INTERNATIONA ;MOLONEY MAURICE M (CA)) 14 October 1993 see page 10, line 23 - line 29	10,14, 15,31
A	EP,A,0 348 348 (CIBA-GEIGY A.-G., SWITZ.) 27 December 1989 see the whole document	19,36
A	EP,A,0 449 375 (GIST BROCADES NV ;MOGEN INT (NL)) 2 October 1991 see the whole document	13
A	EP,A,0 449 376 (GIST BROCADES NV ;MOGEN INT (NL)) 2 October 1991 see the whole document	10,14, 15,31
A	US,A,5 162 601 (SLIGHTOM JERRY L) 10 November 1992 see the whole document	14,15,31
P,X	BIOTECHNOLOGY, vol. 13 , no. 1, January 1995, pages 72-77, XP002005044 VAN ROOIJEN, G.J.H., ET AL.: "Plant seed oil-bodies as carriers for foreign proteins" see the whole document	1-7,10, 11, 20-22, 25-33, 37-45
P,X	PLANT MOLECULAR BIOLOGY, vol. 29, December 1995, pages 1167-1180, XP002005045 PARMENTER, D.L., ET AL.: "Production of biologically active hirudin in plant seeds using oleosin partitioning" see the whole document	1-8, 18-22, 25-30, 32,33, 35-45